Centre de recherche sur les transports (C.R.T.)
Centre for Research on Transportation

CRT-2005-35

INTERMODAL TRANSPORTATION

by

Teodor Gabriel Crainic
Kap Hwan Kim

December 2005

Université de Montréal
C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7 Canada
Téléphone: (514) 343-7575 / Télécopieur: (514) 343-7121

crt@crt.umontreal.ca
www.crt.umontreal.ca

106-01649/

PUBLICATION
CRT-2005-35

INTERMODAL TRANSPORTATION

by

Teodor Gabriel Crainic[1]
Kap Hwan Kim[2]

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), and by the Regional Research Centers Program (Research Center for Logistics Information Technology) granted by the Korean Ministry of Education & Human Resources Development.

[1] Département de management et technologie, Université du Québec à Montréal, and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7 (theo@crt.umontreal.ca)

[2] Department of industrial engineering, Pusan National University, 30 Jangjeon-dong, Geumjeong-gu, Busan 609-735, Republic of Korea (kapkim@pusan.ac.kr)

Dépôt légal – Bibliothèque nationale du Québec
Bibliothèque nationale du Canada, 2005

Abstract

This paper focuses on *Intermodal Freight Transportation*, broadly defined as the transportation of a load from its origin to its destination by a sequence of at least two transportation modes, the transfer from one mode to the next being performed at an intermodal terminal. Container-based transportation services are an important part of intermodal transportation and the backbone of international trade. This importance is reflected in this paper. We also examine issues related to other intermodal systems or components, such as postal services, intermodal rail transportation, and seaport container terminals. Compared to several other application areas, Operation Research models and methods for intermodal freight transportation is still a very young domain. The goal of this paper is thus to be informative and provide a starting point for future research, by providing an overview of the evolution of the field and presenting methodological developments proposed to address a number of important operations and planning issues: system and service design, container fleet management, planning of container terminal operations, and national planning. We focus on models. Algorithmic developments are indicated but not examined in any depth.

Key words: intermodal transportation, freight transportation, operations research.

Résumé

Le transport intermodal de marchandises est vu dans cet article comme une chaîne constituée de plusieurs modes de transport qui interagissent dans des terminaux intermodaux afin d'assurer un service de transport porte à porte. La perspective est celle des institutions qui doivent concevoir le système de transport, ainsi que planifier et opérer ces systèmes. Les problèmes de planification nationale ou régionale sont également traités. Le but de cet article est de présenter une vue d'ensemble du domaine, ainsi que les principaux problèmes et les méthodologies proposées dans la littérature. L'emphase est placée sur les aspects de modélisation ; les développements algorithmiques sont indiqués mais ne sont pas traités en profondeur. Des avenues de recherche sont également identifiées.

Mots-clefs : transport intermodal, transport de marchandises, recherche opérationnelle.

1 Introduction

Intermodal transportation may be defined as the transportation of a person or a load from its origin to its destination by a sequence of at least two transportation modes, the transfer from one mode to the next being performed at an intermodal terminal. The concept is very general and thus, it means many things to many people: transportation of containerized cargo by a combination of truck, rail, and ocean shipping, dedicated rail services to move massive quantities of containers and trailers over long distances, main transportation mode for the international movement of goods, central piece in defining transportation policy for the European Community, trips undertaken by a combination of private (e.g., car) and public (e.g., light rail) transport, and so on. One must therefore start with a few definitions to set the terminology and limit the scope of this paper. First, although both people and freight transportation can be examined from an intermodal perspective, we limit the scope of this chapter to freight.

In one of its most widely accepted meanings, intermodal freight transportations refers to a *multi-modal* chain of *container*-transportation services. This chain usually links the initial shipper to the final consignee of the container (so-called *door-to-door* service) and takes place over long distances. Transportation is often provided by several carriers. In a classical example of an inter-continental intermodal chain, loaded containers leave a shipper's facility by truck either directly to port or to a rail yard from where a train will deliver them to port. A ship will move the containers from this initial port to a port on the other continent, from where they will be delivered to the final destination by a single or a combination of "land" transportation means: truck, rail, coastal or river navigation. Several intermodal terminals are part of this chain: the initial and final sea-port container terminals, where containers are transferred between the ocean navigation and land transportation modes, as well as in-land terminals (rail yards, river ports, etc.) providing transfer facilities between the land modes.

Container transportation is a major component of intermodal transportation and international commerce and this importance is reflected in this chapter. Intermodal transportation is not only about containers and inter-continental exchanges, however. On the one hand, a significant part of international trade that is moved in containers does not involve ocean navigation, land transportation means providing the intermodal chain. On the other hand, other types of cargo may be moved by a chain of transportation means and require intermodal transfer facilities, as illustrated by the definition the European Conference of Ministers of Transport (1993) gives for intermodal transportation: "movement of goods in one and the same loading unit or vehicle, which uses successive, various modes of transportation (road, rail, water) without any handling of the goods themselves during transfers between modes". This last definition is still too restrictive, however. Thus, for example, the transportation of express and regular mail on a regional or national scale is strongly intermodal, using various combinations of road, rail, and air transportation modes, and yet freight is handled (sorted and grouped) in terminals. More generally, the transportation of less-than-vehicle-capacity loads by non-dedicated services is intermodal, since it involves pickup (at origin) and delivery (at destination) operations, usually performed by trucks, at least one long-haul transporta-

tion movement by road, rail, river, or air, as well as transfer activities between these modes in dedicated terminals.

Almost all types of freight carriers and terminal operator may thus be involved in intermodal transportation, either by providing service for part of the transportation chain or by operating an intermodal transportation system (network). We limit the scope of the chapter to the latter with a particular focus on container-based systems, including container terminals in seaports. The national/regional planning perspective, which considers the flow of multiple products on multimodal networks, is also addressed.

Compared to several other application areas, Operations Research models and methods for intermodal freight transportation is still a very young domain. In many cases, there are not, yet, widely accepted models and methodologies. Work is indeed proceeding as this chapter is being written. Therefore, the goal of the chapter is to be informative and provide a starting point for future research. The chapter overviews the evolution of the intermodal-transportation field and presents methodological developments proposed to address a number of important operations and planning issues: system and service design for intermodal transportation networks, container fleet management, container terminal operations and scheduling, national planning. We focus on models. Algorithmic developments are indicated but not examined in any depth.

To structure the presentation, we follow the somewhat classical approach of examining issues, models, and methods according to whether they belong to the *strategic*, *tactic*, or *operational* level of planning and management of operations. The chapter is therefore organized as follows. Section 2 briefly describes freight transportation and the main actors and issues discussed in the chapter. Section 3 is dedicated to system and service network design issues for carriers. Strategic design issues for seaport container terminals are also discussed in this section. Operational-planning issues are addressed in Section 4 with particular emphasis on empty container repositioning. Section 5 is dedicated to tactical and operational planning of operations in container terminals, while Section 6 examines national planning models. Section 7 concludes the chapter with an enumeration of a number of trends and technological developments that may influence intermodal transportation in the future, as well as of promising research directions.

2 Freight Transportation Systems

Demand for freight transportation derives from the interplay between *producers* and *consumers* and the significant distances that often separate them. Producers of goods require transportation services to move raw materials and intermediate products, and to distribute final goods in order to meet customer demands. *Shippers*, which may be the producers of goods or some intermediary firm (e.g., *brokers*), thus generate the *demand* for transportation. *Carriers* answer this demand by *supplying* transportation services. Railways, ocean shipping lines, trucking companies, and postal services are examples of carriers. Considering the type of services they provide, seaports, intermodal platforms, and other such facilities may be described as carriers as well. Governments contribute the infrastructure: roads and highways, as well as significant portions of ports, internal navigation, and rail facilities. Governments also regulate (e.g., dangerous and toxic goods transportation) and tax the industry.

We do not intend to make a detailed presentation of freight transportation. Our goal is rather to describe the basic operation and planning issues for long-haul carriers and terminals that are involved in intermodal transportation, with a particular focus on container-based systems, including container terminals in seaports. The first subsection gives a few statistics and trends relative to container-based transportation. The second and the third are dedicated to long-haul carriers and seaport container terminals, respectively.

Reviews on these issues may be found in Assad (1980), Cordeau, Toth, and Vigo (1998), and Crainic (1988) for rail transportation, Delorme, Roy, and Rousseau (1988) and Powell (1988) for motor carrier transportation, Christiansen, Fagerholt, and Ronen (2004) and Christiansen *et al.* (2006) for maritime transportation, and Günther and Kim (2005) and Steenken, Voß, and Stahlbock (2004) for container port terminals. The reviews proposed by Crainic (2000, 2003), Crainic and Laporte (1997), Daganzo (2005), Dejax and Crainic (1987), Powell (2003), Powell and Topaloglu (2003, 2005), Powell, Jaillet, and Odoni (1995), and Powell, Bouzaïene-Ayari, and Simaõ (2006) are more general in scope, addressing issues relevant for more than one transportation mode or problem. Issues and methodologies related to the pickup and delivery of loads, generally known as vehicle routing and scheduling problems, are not included in this chapter. Interested readers may refer to Golden and Assad (1988), Ball *et al.* (1995), Dror (2000), Toth and Vigo (2002), and Cordeau *et al.* (2006).

2.1 Container Intermodal Transportation

Container-related transportation activities have grown remarkably over the last 10 years and the trend does not show any sign of slowing down as illustrated by the annual world container traffic figures, in millions of TEUs (20 feet equivalent container units), displayed in Table 1 (2005 figures are forecasted; Koh and Kim, 2001; ISL, 2005). The initial impulse to container-based transportation came from the safety it offered regarding loss and damage. Advantages in terms of reduced cargo handling and standardization of transportation and

transfer equipment translate in cost economies and efficient, world-wide door-to-door intermodal service and fuel the growth of the industry. Containerized intermodal transportation supports a significant part of the international movement of goods.

Year	Container traffic (Millions TEU)	Growth rate (%)
1993	113.2	12.5
1995	137.2	9.8
1997	153.5	4.2
1999	203.2	10
2000	225.3	10.9
2001	231.6	2.8
2002	240.6	3.9
2003	254.6	5.8
2004	280.0	10.6
2005	304.0	8.6

Table 1: World Container Traffic

The performance of container-based transportation in international trade has had some remarkable consequences. Ports and container terminals have been built or profoundly modified to accommodate container ships and efficiently perform the loading, unloading, and transfer operations. Container terminal equipment and operating procedures are continuously enhanced to improve productivity and compete, in terms of cost and time, with the other ports to attract ocean-shipping lines. Models for planning container terminal operations are presented in Section 5. Notice that the competitive position of a container port is also dependent on the capacity and efficiency of the land transportation system. The models described in Section 6 may be used for the integrated analysis and planning of port and land transportation systems.

With respect to ocean navigation, efficiency reasons have led to the construction of very large container-carrying ships for inter-continental movements; New container ship capacities are of the order of 8000 to 10000 TEUs. To operate efficiently, such ships must not stop frequently. Moreover, they are too large for the vast majority of ports. Consequently, a new link has been added to the intermodal chain: super-ships stop at a small number of major seaports and containers are transferred to smaller ships for distribution to various smaller ports. Notice that these ships cannot navigate through the Panama Canal. This results in a modification of sea and land shipping routes. The impact of these modifications on particular transportation systems and facilities may be analyzed by using the models of Section 6.

The impact of the growth in containerized trade has been significant on land transportation systems as well. Specialized transportation services have been created, such as the North-American "land-bridges" that provide container transportation by long, double-stack

trains operated by independent subsidiaries of the rail companies between the East and the West coasts and between these ports and the industrial core of the continent (so-called "mini" land-bridges). Dedicated rail services are also being created in Europe to provide shuttle services for container transportation. Section 3.2.3 examines some of the issues associated to planning and operating such dedicated services.

2.2 Customized, Consolidation, and Intermodal Transportation

In an intermodal chain, one may encounter *consolidation* transportation systems, where one vehicle or convoy serves to move freight for different customers with possibly different initial origins and final destinations, and *customized* transportation carriers that provide dedicated service to each particular customer.

Truckload trucking offers a typical example of customized transportation. When the customer calls, the dispatcher assigns to the task a truck and driver (or driving team for very long movements). The truck moves to the customer-designated location, is loaded, and then moves to the specified destination where it is unloaded. The driver then calls the dispatcher to give its position and request a new assignment. The dispatcher may indicate a new load, ask the driver to move empty to a new location where demand should appear in the near future, or have the driver wait and call later.

The truckload carrier thus operates in a highly dynamic environment, where little is known with certainty regarding future demands, waiting delays at customer locations, precise positions of loaded and empty vehicles at later moments in time, and so on. Moreover, the time available to the dispatcher to decide on the next assignment in response to a customer or driver request, is generally very short (most such decisions are performed in real-time). Service is tailored for each customer and the timely assignment of vehicles to profitable demands is very important. The development of efficient *resource management and allocation* strategies is therefore at the heart of the management process. These strategies attempt to maximize the volume of demand satisfied (loads moved) and the associated profits, while making the best use of the available resources: drivers, tractor and trailer fleets, etc. We do not address these issues in this chapter. The interested reader may consult Powell (2003) or Powell, Bouzaïene-Ayari, and Simaõ (2006), for example. Notice, however, that most methodology targeting resource management and allocation issues for customized carriers may also be adapted to address operational-level fleet management issues for consolidation carriers (Section 4).

Ocean navigation services provided by for-hire ships share most of these dynamic and stochastic characteristics. Variations in travel times are in fact larger for sea than for road trips. On the other hand, travel and loading/unloading times are usually much longer, which increases the time available to decide on the next assignment. Railways also provide customized services by dedicating unit trains to individual shippers according. Such decisions respond in most cases to long-term contracts, which implies that demand is known deter-

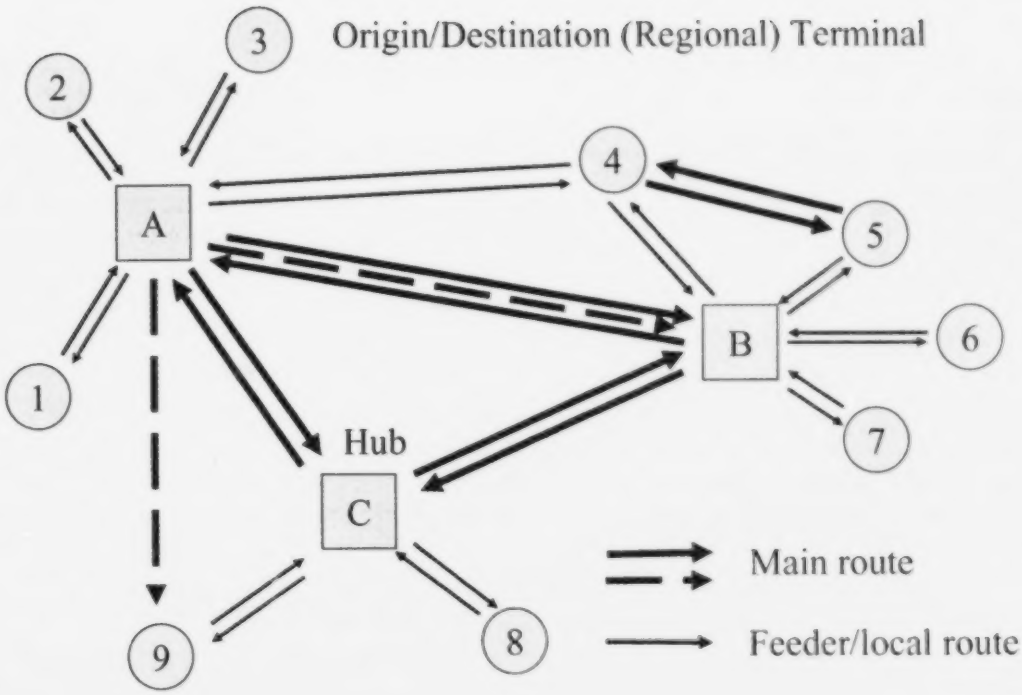

Figure 1: Network with Consolidation Terminals / Hubs

ministically. This contrasts to the stochastic demand that characterizes truckload trucking and for-hire ocean shipping. The setup of such unit trains is therefore part of the regular planning process of operations as described later on in this section.

Customized transportation is not always the appropriate answer to shipper needs. The relations and trade-offs between volume and frequency of shipping, on the one hand, and the cost, frequency, and delivery time of transportation, on the other hand, often dictates the use of *consolidation* transportation services. In passenger transportation, this choice is equivalent, for example, between taking a taxi or using the services of a regular public transport line. Freight consolidation transportation is performed by Less-Than-Truckload (LTL) motor carriers, railways, ocean shipping lines, regular and express postal services, etc. Freight transportation in some countries where a central authority more or less controls a large part of the transportation system also belongs to this category.

Consolidation transportation carriers and fundamentally all intermodal transportation systems are organized as so-called *hub-and-spoke* networks, illustrated in Figure 1. In such systems, service is offered between a certain number of origin-destination points (the local/regional terminals), represented by nodes 1 to 9 in Figure 1. This number is significantly larger than the number of direct, origin to destination services operated by the carrier. Consequently, and to take advantage of economies of scale, low-volume demands are moved first to an intermediate point – a consolidation terminal or a hub (nodes A, B, and C in Figure 1) – such as an airport, seaport container terminal, rail yard, or intermodal platform. At a hub, traffic is consolidated into larger flows that are routed to other hubs by high frequency, high capacity services (thick full and dashed lines in Figure 1). As illustrated in the figure, more than one service, of possibly different modes, may be operated between hubs. Lower frequency services, often operating smaller vehicles, are used between hubs and origin-destination terminals (the regular lines in the figure). When the level of demand justifies it, high frequency, high capacity services may be run between a hub and a regional terminal (links between nodes A and 9) or between two regional terminals (links between nodes 4 and 5). Notice that cargo is brought to and distributed from origin-destination terminals by vehicles performing either pickup and distribution routes or customized services. LTL motor carriers and postal services are examples of the first case, rail and ocean shipping of containers (barges or trucks performing the local transportation activities) of the second. Yet, because the planning of the long-haul activities of consolidation carriers does not generally include these activities, we do not elaborate further on the topic.

A hub-and-spoke organization allows a much higher frequency of service between all origin-destination pairs in the network and a more efficient utilization of resources. The drawback of this type of organization is increased delays due to longer routes and the time spent in terminals. This explains partly why there are very few "pure" hub-and-spoke systems in operation; direct transportation is usually operated for high demand or high priority origin-destination pairs. A second strategy is to separate high and low priority traffic and to dedicate different services and, eventually, infrastructure (e.g., terminals) to each. An example of this trend is the creation of intermodal subdivisions by North-American railways to ensure efficient movements of containers among the major ports and industrial centers of the continent.

To further mitigate the drawbacks of hub-based operations, consolidation carriers engage into rather sophisticate planning activities (Section 3). The carrier operates a series of *services*, each characterized by its own route, stops, frequency, vehicle and convoy type, capacity, speed (travel time), and so on. Internally, services are often collected in an *operational plan* (also referred to as *load* or *transportation* plan), generally accompanied by a *schedule* that indicates departure and arrival times at the terminals of the route. The schedule is partially (e.g., latest delivery of cargo at the origin terminal for on-time delivery at destination) or totally available to customers. The aim of the load plan is to ensure that the proposed services are performed as stated (or as closely as possible), while operating in a rational, efficient, and profitable way. It also indicates how demand is moved though the system using its terminals and transportation services.

Freight demand is defined by its specific origin, destination, and commodity-related physical characteristics (e.g., weight and volume), as well as particular service requirements in terms of delivery conditions, type of vehicle, and so on. A profit or cost is also usually associated to a given demand. Once delivered at the carrier's terminal, the cargo of several customers is sorted, grouped, and loaded into the same vehicle or convoy. It is then moved either directly, when such a service exists, or through a series of services with intermediary operations of transfer and consolidation. More than one consolidation-transfer operation may occur during a trip. In the case of LTL motor carriers, and often for postal services, it is individual loads that are consolidated and loaded into vehicles, trucks, planes, or rail cars. Containerized cargo is not handled before reaching its destination and, thus, consolidation operations involve only the containers, which are loaded into ships, airplanes, or rail cars. This constitutes a first type of consolidation operation: "small" loads into vehicles, containers into ships, etc. In some cases, vehicles are further sorted and consolidated into convoys. The most wise-spread and well-known case is that of railway transportation where, first, cars are grouped into blocks (i.e., the cars in a block travel together, without any re-consolidation operation, until some common, intermediary or final, terminal) and, then, blocks are put together to make up trains. Similar make-up operations, albeit on a smaller scale, also occur for barge trains and multi-trailer assemblies.

Terminals are clearly an important component of consolidation and intermodal transportation systems, and their efficiency is vital to the performance of the entire transport chain. The following sub-section briefly describes main terminal operations and planning issues with a particular focus on container port terminals.

2.3 Consolidation and Intermodal Terminals

Terminals come in several designs and sizes and may be specialized for particular transportation modes and the handling of specific products, or may offer a complete set of services. Major operations performed in consolidation terminals include vehicle loading and unloading, cargo and vehicle sorting and consolidation, convoy make up and break down, and vehicle transfer between services.

When containerized traffic is concerned only the containers are handled, not the cargo they contain. Thus, once loaded on a truck, rail car, barge, or ocean vessel, containers will generally follow the movements and consolidation activities of the respective vehicle, until reaching either the final destination or an intermodal terminal. There are very few such activities for motor carriers and barges. We may mention the transfer of a barge or a trailer from a barge or road train to another. No particular planning model is built for such activities; rather, these are included in the scope of the models used to build the transportation plan of the carrier (Section 3.2).

More complex consolidation operations are performed within railway systems: sorting and consolidation of rail cars into blocks and trains. A rich literature exists on models targeting

these issues and is reviewed in the references indicated at the beginning of the section. In most cases, however, there are no differences between intermodal and regular rail traffic with respect to blocking and train make up planning and operations, even when particular terminals are dedicated to handling intermodal traffic. Research is under way to develop the "next generation" intermodal rail services and terminals, but the efforts dedicated to the associated planning issues are still limited (e.g., Bostel and Dejax, 1998; Macharis and Bontekoning, 2004).

The intermodal transfer of containers between truck and rail, taking place at rail terminals, is specific to intermodal transportation. Containers thus arrive at the rail terminal by truck and are either directly transferred to a rail car or, more frequently, are stacked in a waiting area. Then, containers are picked up from the waiting area and loaded unto rail cars that will be grouped into blocks and trains. The reverse operations take place when containers arrive by train to the terminal and are to be transfered to trucks for their next transport leg. We did not find any paper dedicated to the planning of these operations. The issues are very similar to those arising in container port terminals, however, which are probably the most well-known intermodal transfer facilities. Significant research has been dedicated to the design (Section 3.3) and particularly to the operations (Section 5) of container port terminals. In the following, we briefly describe these operations and planning issues.

The main function of a container port terminal is to provide transfer facilities for containers between sea vessels and land transportation modes, trucks and rail in particular. It is a highly complex system that involves numerous pieces of equipment, operations, and container handling steps (Steenken, Voß, and Stahlbock, 2004). The assignment of resources to tasks and the scheduling of these tasks are thus among the major container port terminal planning issues. Three main areas make up a container terminal. The *sea-side* area encompasses the quays where ships berth and the quay-cranes that provide the loading and unloading of containers into and from ships. The *land-side* area provides the interface with the land transportation system (the so-called hinterland of the port) and encompasses the truck and train receiving gates, the areas where rail cars are loaded and unloaded, and the associated equipment. Trucks are generally loaded and unloaded directly in the *yard area*. This third area is dedicated for the most part to stacking loaded and empty containers for import and export (in some terminals, facilities are also provided for the loading and loading of containers). Various types of yard cranes are associated with this area. So-called *transporters*, primarily yard trucks or automated vehicles, move containers between the three areas. Figure 2 illustrates part of a container port terminal. One ship and three quay cranes are displayed in the sea-side area, while only trucking is shown in the land-side area. Twelve container stacks are displayed in the yard area, as well as one type of yard crane used to transfer containers between yard transporters and outside trucks and stacks, as well as to change the position of containers in the yard as required.

Three main types of handling operations are performed in a container terminal: 1) ship operations associated with berthing, loading, and unloading container ships, 2) receiving/delivery operations for outside trucks and trains, and 3) container handling and storage operations in the yard. When a ship arrives at the container port terminal, it is assigned a

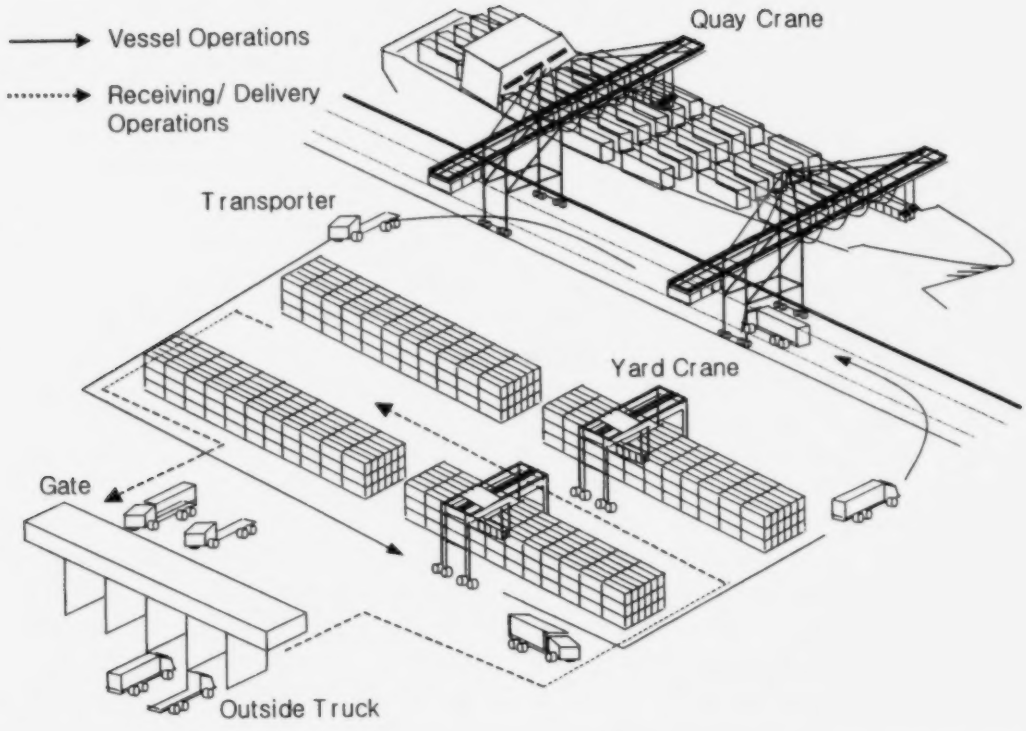

Figure 2: Example of a Container Terminal with an Indirect Transfer System (Park 2003)

berth and a number of quay cranes. Berth space is a very important resource in a container terminal (construction costs to increase capacity are very high, even when space for growth exists) and *berth scheduling* determines the berthing time and position of a container ship at a given quay (Section 5.1). *Quay-crane allocation* is the process of determining the vessel that each quay crane will serve and the associate service time (Section 5.2). *Stowage sequencing* determines the sequence of unloading and loading containers, as well as the precise position each container being loaded into the ship is to be placed (Section 5.3). During the unloading operation, a quay crane transfers a container from a ship to a transporter. Then, the transporter delivers the import (unloading) container to a yard crane that picks it up and stacks it into a given position in the yard. This sequence of operations is called *indirect transfer*. Some terminals use a *direct transfer* system where the equipment used to move containers between the quay and the yard will also stack them. For export (loading) operation, the process is carried out in the opposite direction.

On the land-side, the *receiving* and *delivery* operations provide the interface between the container terminal activities and the external movements. A receiving operation starts when containers arrive at the gate of the terminal carried by one or several outside trucks or a train. Containers are inspected at the gate to check for damages (to the container not its content) and whether all documents are in order. Also at the gate, information regarding where the container is to be stored is provided to the truck driver. When the outside truck arrives at the indicated transfer point, a yard crane lifts a container from the truck and

stacks it according to the plan. When containers arrive by rail, the rail cars are brought in the rail area where containers and documents are examined. Containers are then transferred by a gantry crane to a transporter, which delivers them to the yard and stacks them. In the case of a delivery operation, the yard equipment delivers a container onto an outside truck, which leaves the port, or onto a transporter which delivers the container to the rail area and loads it onto the designated car.

The sea and land-side operations interact with the yard container handling and storage operation through the information on where the containers are or must be stacked within the yard. How containers are stored in the yard is one of the important factors that affect the turn-around time of ships and land vehicles. The *space-allocation* problem is concerned with determining storage locations for containers either individually or as a group. Yard storage space is pre-assigned to containers of each ship arriving in the near future to maximize the productivity of the loading and unloading operations (Section 5.4).

A container yard consists of blocks of containers, which are separated by aisles for transporters as shown in Figure 2. A block consists of 25 to 35 yard bays, and a yard bay has 6 to 10 stacks of containers. Container handling and storage operations include the management and handling of containers while they are in storage in the yard and thus occur between the receiving and delivery operations and the ship operations. Container-handling equipment performs the placement of containers into storage and their retrieval when needed. Yard cranes move along blocks of containers to yard bays to perform these operations. Planning these operations is part of the *equipment-assignment* process, which allocates tasks to container-handling equipment. Based on the quay-crane schedule, one or two yard cranes are assigned to each quay crane for loading and unloading. The remaining yard cranes are allocated to receiving and delivery operations. Terminal operators aim to assign and operate yard cranes in such a way that inefficient moves and interferences among yard cranes are minimized (Section 5.5).

3 System and Service Network Design

This section is dedicated to models aimed at strategic and tactical planning issues for freight carriers and seaport container terminals. At the strategic level, we focus on decisions that concern the design of the physical infrastructure network: where to locate terminals (e.g., consolidation terminals, rail yards, intermodal platforms, and so on)? what type and quantity of equipment (e.g., cranes) to install at each facility? what type of lines or capacity to add? what lines or facilities to abandon? which customer zones to serve directly and how to serve the others? and so on. We group these issues under the label *system design*. The term *service network design* covers tactical planning issues for consolidation transportation firms: on what routes to provide service? what type of service (mode) to use? how often to offer service on each route and according to what schedule? how to route the loads through the physical and service networks? how to distribute the work among the terminals of the system? The output of the process is a transportation (load) plan. The goals are customer satisfaction and cost-efficient utilization of resources (assets) leading to profits.

Most freight carriers face system and service network design issues, irrespective of their involvement in intermodal transportation activities. It is not our intention, however, to present a comprehensive and detailed treatment of the subject. We focus rather on the main issues and modeling efforts that bear directly on intermodal transportation.

3.1 Models for System Design

The literature on system design models for freight transportation is not rich. Such issues are often addressed by evaluating alternatives using network models for tactical or operational planning of transportation activities. When formal models are proposed, they generally take the form of discrete location formulations to address issues related to the location of consolidation or hub terminals and the routing of demand from its origin to its destination terminals. The routing of flows determines the direct connections (physical infrastructure or service links) between origins, destinations, and consolidation terminals. When these connections must be explicitly decided (e.g., the allocation of "local" terminals to major classification facilities), a combined location-network-design formulation is often used. All formulations aim to capture the potential economies of scale associated with the consolidation of freight.

An extensive literature exists on location and network design models and solution methods: Mirchandani and Francis (1990), Daskin (1995), Drezner (1995), Labbé, Peeters, and Thisse (1995), Labbé and Louveaux (1997), Crainic and Laporte (1997), Drezner and Hamacher (2002), and Daskin and Owen (2003) review location issues and literature, while Magnanti and Wong (1984), Minoux (1989), Nemhauser and Wolsey (1988), Salkin and Mathur (1989), Ahuja, Magnanti, and Orlin (1993), Balakrishnan, Magnanti, and Mirchandani (1997), and Crainic (2000) survey the network design field.

In the following, we focus on the contributions related to intermodal transportation, the location of consolidation facilities in particular. All the models presented are static and deterministic location formulations. They assume that all problem components, particularly the demand as well as the cost and profit structure, will not vary during the planning period for which their evaluation is performed, usually from six to twelve months. Moreover, problems where the design decisions have long-term effects (e.g., location of hubs) also assume that variances will be negligible for the foreseeable future. These observations emphasize the need for research into time-dependent stochastic models for system design problems. Most problems presented in this subsection are NP-Hard and thus heuristics are the solution method of choice in almost all cases.

3.1.1 Location with balancing requirements

The multicommodity location problem with balancing requirements was first introduced by Crainic, Dejax, and Delorme (1989) in the context of the management of a heterogeneous fleet of containers by an international maritime shipping company. The land operations of the carrier proceed as follows. Once a ship arrives at port, the company has to deliver loaded containers, which may come in several types and sizes, to designated in-land destinations. Following their unloading by the importing customer, empty containers are moved to a depot. From there, empty containers may be delivered to customers which request them for subsequent shipping of their own products. In addition, empty containers are often moved, or *repositioned*, to other depots. These interdepot movements are a consequence of regional imbalances in empty container availabilities and needs throughout the network: some areas lack containers of certain types, while others have surpluses. These balancing movements of empty containers among depots differentiate this problem from classical location-allocation applications. Interdepot movements are performed by high density transport (rail, typically) and their unit transportation cost is lower than for the other types of movements. The general problem is therefore to locate depots and allocate customers to depots (for each type of container and direction of movement) in order to collect the supply of empty containers available at customers' sites and to satisfy the customer requests for empty containers, while minimizing the total operating costs: the costs of opening and operating the depots, and the costs generated by customer-depot and interdepot movements.

The formulation proposed by Crainic, Dejax, and Delorme (1989) is based on a directed network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is the set of nodes or vertices and $\mathcal{A}$ is the set of arcs or links. The set of nodes is partitioned into three subsets: $\mathcal{O}$, the set of origin nodes (supply customers); $\mathcal{D}$, the set of destination nodes (demand customers); and $\mathcal{H}$, the set of hubs or consolidation nodes (depots). The sets of customers adjacent to each depot $j \in \mathcal{H}$ are identified as $\mathcal{O}(j) = \{i \in \mathcal{O} : (i,j) \in A\}$ and $\mathcal{D}(j) = \{i \in \mathcal{D} : (j,i) \in A\}$. It is assumed that there exists at least one origin or destination adjacent to each depot j $(\mathcal{O}(j) \cup \mathcal{D}(j) \neq \emptyset)$. The sets of depots adjacent to each node $i \in N$ in both directions are defined as $\mathcal{H}^+(i) = \{j \in \mathcal{H} : (i,j) \in A\}$ and $\mathcal{H}^-(i) = \{j \in \mathcal{H} : (j,i) \in A\}$. The problem description does not allow direct customer-to-customer movements of containers. Hence, the set of arcs $\mathcal{A}$ may be

partitioned into three subsets: customer-to-depot arcs, $A_{OH} = \{(i,j) \in A : i \in \mathcal{O},\ j \in \mathcal{H}\}$; depot-to-customer arcs, $A_{HD} = \{(i,j) \in A : i \in \mathcal{H},\ j \in \mathcal{D}\}$; and depot-to-depot arcs, $A_{HH} = \{(i,j) \in A : i \in \mathcal{H},\ j \in \mathcal{H}\}$. The commodities (types of containers) that move through the network are represented by the set $\mathcal{P}$.

For each supply customer $i \in \mathcal{O}$, the supply of commodity p is noted $o_i^p \geq 0$, while for each demand customer $i \in \mathcal{D}$, the demand for commodity p is noted $d_k^p \geq 0$. A non-negative cost c_{ij}^p is incurred for each unit of flow of commodity p moving on arc (i,j). In addition, for each depot $j \in \mathcal{H}$, a non-negative fixed cost f_j is incurred if the depot is opened.

Let x_{ij}^p represent the flow of commodity p moving on arc (i,j), and y_j the binary location variable that takes value 1 if depot j is opened, and value 0 otherwise. The problem is then formulated as:

$$\text{Minimize} \quad \sum_{j \in \mathcal{H}} f_j y_j + \sum_{p \in \mathcal{P}} \left(\sum_{(i,j) \in A_{OH}} c_{ij}^p x_{ij}^p + \sum_{(j,i) \in A_{HD}} c_{ji}^p x_{ji}^p + \sum_{(j,k) \in A_{HH}} c_{jk}^p x_{jk}^p \right) \tag{1}$$

$$\text{Subject to} \quad \sum_{j \in \mathcal{H}^+(i)} x_{ij}^p = o_i^p \qquad i \in \mathcal{O},\ p \in \mathcal{P} \tag{2}$$

$$\sum_{j \in \mathcal{H}^-(i)} x_{ji}^p = d_k^p \qquad i \in \mathcal{D},\ p \in \mathcal{P} \tag{3}$$

$$\sum_{i \in \mathcal{D}(j)} x_{ji}^p + \sum_{k \in \mathcal{H}^+(j)} x_{jk}^p - \sum_{i \in \mathcal{O}(j)} x_{ij}^p - \sum_{k \in \mathcal{H}^-(j)} x_{kj}^p = 0 \qquad j \in \mathcal{H},\ p \in \mathcal{P} \tag{4}$$

$$x_{ij}^p \leq o_i^p y_j \quad j \in \mathcal{H},\ i \in \mathcal{O}(j),\ p \in \mathcal{P} \tag{5}$$

$$x_{ji}^p \leq d_k^p y_j \quad j \in \mathcal{H},\ i \in \mathcal{D}(j),\ p \in \mathcal{P} \tag{6}$$

$$x_{ij}^p \geq 0 \qquad (i,j) \in A,\ p \in \mathcal{P} \tag{7}$$

$$y_j \in \{0,1\} \qquad j \in \mathcal{H} \tag{8}$$

Constraints (2) and (3) ensure that supply and demand requirements are met, relations (4) correspond to flow conservation constraints at depot sites, while equations (5) and (6) forbid customer-related movements through closed depots. The analogous constraints for the interdepot flows are redundant if costs satisfy the triangle inequality and are not included in the formulation. The problem has been addressed by dual-ascent heuristics (Crainic and Delorme 1995), sequential and parallel branch-and-bound (Crainic, Delorme, and Dejax 1993; Gendron and Crainic 1995, 1997; Bourbeau, Gendron, and Crainic 2000), sequential and parallel tabu search (Crainic *et al.* 1993; Crainic, Toulouse, and Gendreau 1995a,b, 1997; Gendron, Potvin, and Soriano 1999). Gendron, Potvin, and Soriano (2003a,b have studied the capacitated version of the problem and proposed sequential and parallel meta-heuristics.

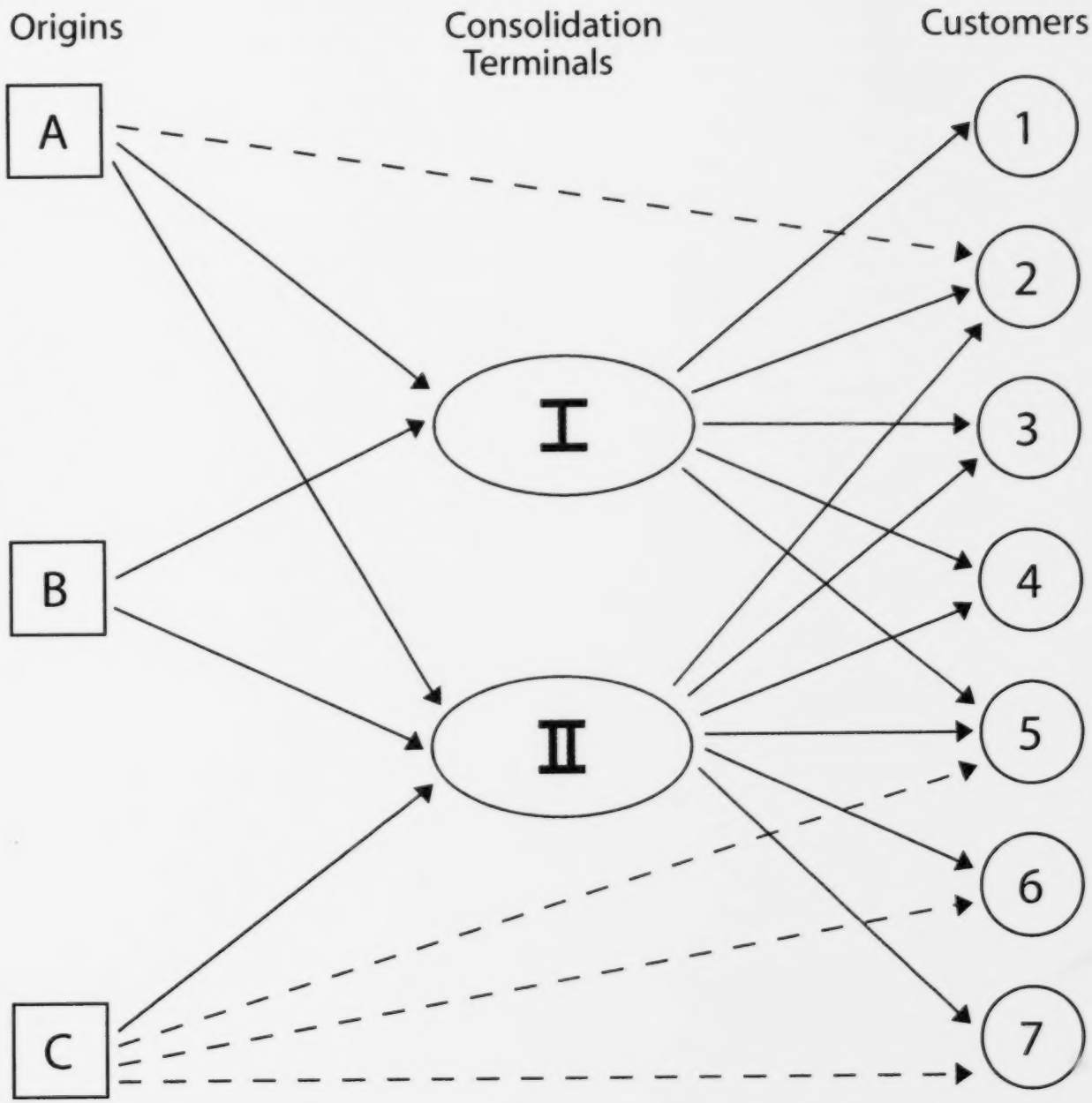

Figure 3: Production-Distribution Network

3.1.2 Multicommodity Production-Distribution

The multicommodity production-distribution problem is a simplified version of the previous problem where no inter-hub movements are considered. In this case, commodities $p \in \mathcal{P}$ may be shipped from their origins $i \in \mathcal{O}$ to their destinations $k \in \mathcal{D}$ either directly or via a consolidation terminal $j \in \mathcal{H}$. (Note that sets $\mathcal{O}, \mathcal{D}$, and $\mathcal{H}$ are not necessarily disjoint.) The main decisions addressed by the production-distribution formulations are the number and location of consolidation terminals and the product flow pattern through the system, either directly from the origin to destination or through a consolidation terminal. Particularly relevant for intermodal transportation system design is the representation of the economies of scale in transportation costs associated with concentrated flows (e.g., Croxton, Gendron, and Magnanti, 2003, 2004). Figure 3 illustrates these concepts for 3 origins, seven destinations, and two possible locations for consolidation centers. Plain arrows indicate the corresponding movements. Dashed arrows represent the direct shipments that are permitted.

The capacity of a consolidation terminal located at site $j \in \mathcal{H}$ is denoted u_j. The transportation cost per unit of flow of commodity p from origin i to destination k transiting through consolidation terminal j is denoted c^p_{ijk}, while c^p_{ik} stands for the unit transportation cost of direct movements. The corresponding flow routing decision variables are denoted x^p_{ijk} and x^p_{ik}, respectively. All other notation is the same as in the previous problem. The model is then:

$$\text{Minimize} \quad \sum_{j\in\mathcal{H}} f_j y_j + \sum_{p\in\mathcal{P}}\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{H}}\left(c^p_{ik}x^p_{ik} + \sum_{k\in\mathcal{D}} c^p_{ijk}x^p_{ijk}\right) \tag{9}$$

$$\text{Subject to} \quad x^p_{ijk} \leq min\{o^p_i, u_j, d^p_k\}y_j \qquad i \in \mathcal{O},\ j \in \mathcal{H}\ , k \in \mathcal{D},\ p \in \mathcal{P} \tag{10}$$

$$\sum_{p\in\mathcal{P}}\sum_{i\in\mathcal{O}}\sum_{k\in\mathcal{D}} x^p_{ijk} \leq u_j y_j \qquad j \in \mathcal{H} \tag{11}$$

$$\sum_{k\in\mathcal{D}}\left(x^p_{ik} + \sum_{j\in\mathcal{H}} x^p_{ijk}\right) \leq o^p_i \qquad i \in \mathcal{O},\ p \in \mathcal{P} \tag{12}$$

$$\sum_{i\in\mathcal{O}}\left(x^p_{ik} + \sum_{j\in\mathcal{H}} x^p_{ijk}\right) = d^p_k \qquad k \in \mathcal{D},\ p \in \mathcal{P} \tag{13}$$

$$y_j = 0 \text{ or } 1 \qquad j \in \mathcal{H} \tag{14}$$

$$x^p_{ijk} \geq 0\ , x^p_{ik} \geq 0 \qquad i \in \mathcal{O}\ , k \in \mathcal{D},\ j \in \mathcal{H},\ p \in \mathcal{P}. \tag{15}$$

In this formulation, expression (9) minimizes the total cost of opening the consolidation centers and distributing the commodities. Constraints (10) and (11) ensure that only open terminals are used. Relation (11) also enforces the consolidation-terminal capacity constraint. Commodity supplies and demands are enforced by relations (12) and (13), respectively. When demand is specified by origin-to-destination pairs

d_{ik}^p: Quantity of commodity p to be transported from origin terminal i to destination terminal k,

the previous formulation is modified by replacing relations (10) with (16) and constraints (12) and (13) with (17):

$$x_{ijk}^p \leq min\{u_j, d_{ik}^p\} y_j \quad i \in \mathcal{O},\ j \in \mathcal{H}\ , k \in \mathcal{D},\ p \in \mathcal{P} \tag{16}$$

$$x_{ik}^p + \sum_{j \in \mathcal{H}} x_{ijk}^p = d_{ik}^p \qquad i \in \mathcal{O},\ k \in \mathcal{D},\ p \in \mathcal{P}. \tag{17}$$

Many variants of these formulations may be found in the location literature dealing with practical considerations such as partial, commodity-specific capacities at terminals or handling costs at consolidation terminals (e.g., Aikens, 1985). Piecewise linear functions are often used to represent economies of scale in transportation costs associated with concentrated flows (e.g., Croxton, Gendron, and Magnanti, 2003, 2004). Klincewicz (1990) proposed such a formulation for an uncapacited problem with single-sourcing for each commodity and no direct origin-to-destination movements. When linear transportation costs are assumed either for the origin to consolidation center movements or from the latter to destination terminals, the problem decomposes into concave-cost uncapacitated location problems that can be formulated as uncapacitated plant location problems (with one facility for each piecewise linear segment) for which efficient solution methods exist (e.g., Erlenkotter, 1978). Heuristics that solve a series of these formulations are proposed for the general case. Heuristics combining Lagrangian relaxations and ad-hoc rules based on the logistics characteristics of the system (costs and distances, principally) have also been proposed for the capacitated version of the problem (9)-(15) (Pirkul and Jayaraman, 1996, 1998).

3.1.3 Hub-Location Models

Consolidation transportation systems are generally structured as *hub-and-spoke* networks (Section 2 and Figure 1). Hub-location models may be used when one must decide simultaneously the location of the hubs (consolidation terminals) and the allocation of regional terminals or customers to the hubs. The location with balancing requirements and the multicommodity production-distribution problems presented in the previous subsections are particular hub-location cases where the allocation of regional terminals or customers to potential hubs does not involve any setup (fixed) cost or has already been decided, respectively. Hub-location formulations appear in many application areas (Campbell, 1994b) and several variants have been proposed and studied (Campbell, Ernst, and Krishnamoorthy, 2002; Ebery *et al.*, 2000).

The basic model assumes that all traffic passes through two hubs on its route from its origin to its destination, no hub capacities, no direct transport between non-hub terminals, and no fixed costs for establishing a link between a regional and a consolidation terminal.

In all hub-based problems, inter-hub transportation is assumed to be more efficient due to the concentration of flows. Consequently, inter-hub links are assigned a lower unit cost than links representing the other movements in the system. Most of the notation is similar to that of the previous models. Let us add the following decision-variable definitions:

$y_j =$ 1, if a consolidation terminal is located at site j and 0, otherwise;
$y_{ij} =$ 1, if terminal i is linked to hub j and 0, otherwise;
x^p_{ijlk} : Flow of commodity p with origin i and destination k that passes through hubs j and l, in that order.

The following formulation assumes that exactly P hubs have to be located out of the $|\mathcal{H}|$ potential sites:

$$\text{Minimize} \quad \sum_{p\in\mathcal{P}}\left\{\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{H}} c^p_{ij}y_{ij}\left(\sum_{l\in\mathcal{H}}\sum_{k\in\mathcal{D}} x^p_{ijlk}\right)+\sum_{l\in\mathcal{H}}\sum_{k\in\mathcal{D}} c^p_{lk}y_{lk}\left(\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{H}} x^p_{ijlk}\right)\right.$$
$$\left.+\sum_{j\in\mathcal{H}}\sum_{l\in\mathcal{H}} c^p_{jl}y_{ij}y_{lk}\left(\sum_{i\in\mathcal{O}}\sum_{k\in\mathcal{D}} x^p_{ijlk}\right)\right\} \tag{18}$$

$$\text{subject to} \quad \sum_{j\in\mathcal{H}} y_j = P \tag{19}$$

$$\sum_{j\in\mathcal{H}}\sum_{l\in\mathcal{H}} x^p_{ijlk} = d^p_{ik} \qquad i\in\mathcal{O},\ k\in\mathcal{D},\ p\in\mathcal{P} \tag{20}$$

$$y_{ij} \le y_j \qquad i\in\mathcal{O},\ j\in\mathcal{H} \tag{21}$$

$$y_{lk} \le y_l \qquad k\in\mathcal{D},\ l\in\mathcal{H} \tag{22}$$

$$x^p_{ijlk} \le d^p_{ik}y_j \qquad i\in\mathcal{O},\ k\in\mathcal{D},\ j,l\in\mathcal{H},\ p\in\mathcal{P} \tag{23}$$

$$x^p_{ijlk} \le d^p_{ik}y_l \qquad i\in\mathcal{O},\ k\in\mathcal{D},\ j,l\in\mathcal{H},\ p\in\mathcal{P} \tag{24}$$

$$y_j = 0 \text{ or } 1 \qquad j\in\mathcal{H} \tag{25}$$

$$y_{ij} = 0 \text{ or } 1 \qquad i\in\mathcal{O},\ j\in\mathcal{H} \tag{26}$$

$$x^p_{ijlk} \ge 0 \qquad i\in\mathcal{O},\ k\in\mathcal{D},\ j,l\in\mathcal{H},\ p\in\mathcal{P}. \tag{27}$$

The objecti function (18) minimizes the total transportation cost of the system. Due to its similarity with the *p-median location* problem, this class of formulations is called the *p-hub median* problem (Campbell, 1994a). Note, however, that despite the name similarity, several properties of the p-median problem do not hold for the p-hub problem (e.g., the assignment of demand nodes to the closest open facility is not optimal for p-hub median problems).

Formulation (18)-(27) is quadratic and was first introduced by O'Kelly (1987). Campbell (1994a) introduced the first linearization mechanism, later refined by Campbell (1996), Skorin-Kapov, Skorin-Kapov, and O'Kelly (1996), and O'Kelly *et al.* (1996). The mechanism yields a *path*-based, linear mixed-integer formulation:

$$\text{Minimize} \quad \sum_{p\in\mathcal{P}}\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{H}}\sum_{l\in\mathcal{H}}\sum_{k\in\mathcal{D}} c^p_{ijlk}x^p_{ijlk} \tag{28}$$

subject to (19), (20), (23) - (25), and (27), where

c^p_{ijlk}: Unit cost of transportation for commodity p from origin i to destination k with consolidation at hubs j and l, in that order.

The two preceding formulations allow multiple allocations of origin and destination terminals to hubs. To enforce single allocation of terminals to hubs, constraints (23) and (24) are replaced in the first formulation by

$$\sum_{j \in \mathcal{H}} y_{ij} = 1 \qquad i \in \mathcal{O} \tag{29}$$

$$\sum_{l \in \mathcal{H}} y_{lk} = 1 \qquad k \in \mathcal{D} \tag{30}$$

$$x^p_{ijlk}/d^p_{ik} = y_{ij} \quad i \in \mathcal{O},\ k \in \mathcal{D},\ j,l \in \mathcal{H},\ p \in \mathcal{P} \tag{31}$$

$$x^p_{ijlk}/d^p_{ik} = y_{lk} \quad i \in \mathcal{O},\ k \in \mathcal{D},\ j,l \in \mathcal{H},\ p \in \mathcal{P} \tag{32}$$

Both these constraints and the y_{ij}, y_{lk} variables must be added to the linearized formulation to enforce single terminal allocation.

Both hub-location formulations are difficult to solve. Heuristics have thus been mainly proposed (e.g., O'Kelly, 1987; Aykin, 1990; Klincewicz, 1991; Campbell, 1996). Meta-heuristics have led to improved results: Neural Networks (Smith, Krishnamoorthy, and Palaniswami, 1996), Simulated Annealing (Ernst and Krishnamoorthy, 1996), and Tabu Search (Klincewicz, 1992 and Skorin-Kapov and Skorin-Kapov, 1994). Lower bounds have also been proposed (e.g., O'Kelly, 1992a; O'Kelly, Skorin-Kapov, and Skorin-Kapov, 1995; Skorin-Kapov, Skorin-Kapov, and O'Kelly, 1996), as well as a number of methods that find the optimal solution to problem instances of limited size (Skorin-Kapov, Skorin-Kapov, and O'Kelly, 1996; O'Kelly *et al.*, 1996; Ernst and Krishnamoorthy, 1996).

A number of important extensions to these models may be formulated. Fixed cost terms $\sum_{j \in \mathcal{H}} f_j y_j$ added to the objective function (and dropping constraint (19) on the number of hubs to open) may be used to represent the costs of opening consolidation centers (O'Kelly, 1992b). Klincewicz (1996) proposed dual-based heuristics for this formulation, while Abdinnour-Helm and Venkataramanan (1998) developed a genetic search meta-heuristic and a branch-and-bound algorithm. Limited-sized problem instances have been addressed.

Additional extensions consider fixed costs for allocating origin and destination terminals to hubs, hub capacities, more complex routing patterns (e.g., more than two hubs), minimum levels of traffic in order to allow a terminal-to-hub connection, etc., and are better able to capture the complexity of transportation systems. Such hub network design formulations are very difficult to solve, however, and relatively little research had been conducted (e.g., Aykin, 1994, 1995a,b; Kuby and Gray, 1993; Jaillet, Song, Yu, 1996). Significant research is required in this domain.

3.2 Service Network Design

Service network design formulations are used to build a transportation (or load) plan to ensure that the system operates efficiently, serves demand, and ensures the profitability of the firm. The physical infrastructure (e.g., the terminal locations) and the available resources are fixed for these problems. Service network design problems address the system-wide planning of operations to decide the selection, routing, and scheduling of services, the consolidation activities in terminals, and the routing of freight of each particular demand through the physical and service network of the company. The goal is cost-efficient operation together with timely and reliable delivery of demand according to customer specifications and the targets of the carrier.

Service network design problems are difficult due to the strong interactions among system components and decisions and the corresponding tradeoffs between operating costs and service levels. To illustrate, consider the routing of a shipment between two terminals of the intermodal transportation system illustrated in Figure 1. (Assume, for simplicity, that each link corresponds to a service.) A shipment that originates at terminal 4 with destination terminal 8 may be routed according to a number of strategies, including 1) Direct through its designated hub B; 2) Indirect through hub A or terminal 5 and then through B; 3) Indirect through hubs A and C. The number of strategies increases rapidly when one considers additional factors such as the type of operation (e.g., consolidation or simple cargo transfer from one service to another) or consolidation policy.

Which alternative is the "best"? Each has its own cost and time measures that result from the characteristics of each terminal and service, as well as from the routing of all other shipments. Thus, for example, strategies based on re-consolidation and routing through intermediate terminals may be more efficient when direct services are offered rarely due to low levels of traffic demand. Such strategies would probably result in higher equipment utilization and lower waiting times at the original terminal; hence, in a more rapid service for the customer. The same decision would also result, however, in additional unloading, consolidation, and loading operations, creating larger delays and higher congestion levels at terminals, as well as a decrease in the total reliability of the shipment. Alternatively, offering direct service (introducing the service or increasing its frequency) would imply faster and more reliable service for the corresponding traffic and a decrease in the level of congestion at some terminals, but at the expense of additional resources, thus increasing the direct costs of the system.

Service network design thus integrates two types of major decisions. The first is to determine the service network, that is, to *select* the routes – origin and destination terminals, physical route and intermediate stops – on which services will be offered and the characteristics of each service: mode, *frequency* or *schedule*, etc. The second major type of decision is to determine the routing of demand, that is, the *itineraries* used to move the flow of each demand: services and terminals used, operations performed in these terminals, etc. The service network specifies the movements through space and time of the vehicles and convoys of the

various modes considered. Operating rules specifying, for example, how cargo and vehicles may be sorted and consolidated are sometimes specified at particular terminals and become part of the service network (this is the case, in particular, for rail carriers). The itineraries used to move freight from origins to destinations determine the flows on the services and through the terminals of the service network.

Minimization of the total operating costs is the primary optimization criterion, reflecting the traditional objectives of freight carriers and expectation of customers to "get there fast at lowest possible cost". Increasingly, however, customers not only expect low rates, but also require a high-quality service, measured by speed, flexibility, and reliability. Service performance measures modeled, in most cases, by delays incurred by freight and vehicles or by the respect of predefined performance targets are then added to the objective function of the network optimization formulation. The resulting generalized cost function thus captures the tradeoffs between operating costs and service quality.

Several efforts have been directed toward the formulation of service network design models. Reviews are presented by Assad (1980), Crainic (1988, 2000, 2003), Delorme, Roy, and Rousseau (1988), and Cordeau, Toth, and Vigo (1998). Most of these contributions aim specific carrier types and modes of transportation (rail, less-than-truckload trucking, navigation, etc.) and are thus covered in more depth in other chapters of this book. In the following, we briefly present basic modeling approaches and examine a number of contributions aimed at intermodal transportation.

One may distinguish between static and time-dependent service network design formulations. The former assume that demand does not vary during the planning period that is considered. The time dimension of the service network is then implicitly considered through the definition of services and inter-service operations at terminals. Time-dependent formulations include an explicit representation of movements in time and usually target the planning of *schedules* to support decisions related to *when* services depart. Most service network design models proposed in the literature take the form of *deterministic, fixed cost, capacitated, multicommodity network design* formulations.

3.2.1 Static Service Network Design

Service network design assumes a physical network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ on which the carrier may move its vehicles. On this network, the carrier operates (or makes use of) a number of terminals, with various characteristics and functionalities (loading/unloading stations, regional terminals with limited sorting capabilities, consolidation hubs, etc.). Let nodes in $\mathcal{H} \subseteq \mathcal{N}$ correspond to locations where the terminals are situated and assume, for simplicity, that all terminals can perform all operations.

The *service network* specifies the transportation services that could be offered. Each *service* $s \in \mathcal{S}$ is characterized by its 1) mode, which may represent either a specific trans-

portation mode (e.g., rail and truck services may belong to the same service network), or a particular combination of traction and service type (e.g., fast navigation lines by large container ships providing shuttle service between South-East Asia and the west coast of North America); 2) route, defined as a path in $\mathcal{A}$, from its origin terminal to its destination terminal, with intermediary terminals where the service stops and work may be performed (e.g., a navigation line that stops at pre-defined ports for unloading and loading containers); 3) capacity, which may be measured in load weight or volume, number of containers, number of vehicles (when convoys are used to move several vehicles simultaneously), or a combination thereof; 4) service class that indicates characteristics such as preferred traffic or restrictions, speed and priority, etc. To design the service network thus means to decide what service to include in the transportation plan such that the demands and the objectives of the carrier are satisfied. When a service may be operated repeatedly during the planning period (e.g., several similar ships departing during the month and visiting the same ports in the same order), the design must also determine the frequency of each service.

Crainic and Rousseau (1986) proposed a multimodal multicommodity path-flow service network design formulation. In their model, a commodity $p \in \mathcal{P}$ is defined as a triplet (origin, destination, type of product (or vehicle)) and traffic moves according to *itineraries*. An itinerary $l \in \mathcal{L}^p$ for commodity p specifies the service path used to move (part of) the corresponding demand: the origin and destination terminals, the intermediary terminals where operations (e.g., consolidation and transfer) are to be performed, and the sequence of services between each pair of consecutive terminals where work is performed. The demand of product p is denoted d_p. Flow routing decisions are then represented by decision variables h_l^p indicating the volume of product p moved by using its itinerary $l \in \mathcal{L}^p$. *Service frequency* decision variables y_s, $s \in \mathcal{S}$, define the level of service offered, i.e., how often each service is run during the planning period.

Let $y = \{y_s\}$ and $h = \{h_l^p\}$ be the decision-variable vectors. The model minimizes the total generalized system cost, while satisfying the demand for transportation and the service standards:

$$\text{Minimize} \quad \sum_{s \in \mathcal{S}} \mathcal{F}_s(y) + \sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}^p} \mathcal{C}_l^p(y,h) + \Theta(y,h) \tag{33}$$

$$\text{subject to} \quad \sum_{l \in \mathcal{L}^p} h_l^p = d_p \qquad p \in \mathcal{P} \tag{34}$$

$$y_s \geq 0 \text{ and integer} \qquad s \in \mathcal{S} \tag{35}$$

$$h_l^p \geq 0 \qquad l \in \mathcal{L},\ p \in \mathcal{P} \tag{36}$$

where,

$\mathcal{F}_s(y)$: Total cost of operating service s;

$\mathcal{C}_l^p(y,h)$: Total cost of moving (part of) product p demand by using its itinerary l;

$\Theta(y,h)$: Penalty terms capturing various relations and restrictions, such as the limited service capacity;

The objective function defines the total system cost and includes the total cost of operating a service network at given frequencies, the total cost of moving freight by using the selected itineraries for each commodity, as well as a number of terms translating operational and service restrictions, such as facility and service capacities and on-time delivery targets, into monetary values. This third term models, for example, service capacity restrictions as utilization targets, over-assignment of traffic being permitted at the expense of additional costs and delays. The introduction of the third term aims to enhance the capability of the model to identify tradeoffs between the cost of increasing the level of service (service frequency) and the cost of routing freight on less interesting itineraries.

The objective function computes a generalized cost, in the sense that it may include various productivity measures related to terminal and transportation operations. In addition to the actual costs of performing the operations, one may thus explicitly consider the costs, delays, and other performance measures related to the quality and reliability of the service offered. Non-linear congestion functions were thus used to reflect the increasingly larger delays that result when facilities of limited capacity (e.g., consolidation terminals) must serve a growing volume of traffic brought by vehicles of different services carrying freight for different products on various itineraries. The notation aims to convey this utilization of more general functional forms.

The model was adapted and applied to rail, long-haul LTL trucking (railway transportation of trailers was included as a possible mode), and express letter services (e.g., Crainic, Ferland, and Rousseau 1984; Roy and Delorme, 1989; Crainic and Roy 1988, and Roy and Crainic (1992). The original solution method described by Crainic and Rousseau (1986) combines a heuristic that iteratively decreases frequencies from initial high values, and a convex network optimization procedure to distribute the freight. The latter makes use of column generation to create itineraries and descent procedures to optimize the flow distribution.

Most service network design models proposed in the literature use arc-flow-based formulations and focus on the selection decisions. This results into $\{0, 1\}$ network design formulations (e.g., the liner service design models presented by Christiansen *et al.* 2006). Minimum thresholds are imposed in some cases on the volume of traffic required to allow the selection of a transportation service (e.g., the load planning model for LTL motor carriers introduced by Powell, 1986, Powell and Sheffi 1983, 1986, 1989, and Braklow *et al.*, 1992).

3.2.2 Design of Postal Services

The design of postal services, regular and those dedicated to express letter and package delivery, yields particularly large and complex service network design formulations. These applications involve air and land long-haul transportation, as well as local pick up and delivery operations. Land long-haul transportation is often performed by using trucks, although rail is also used in some countries. It should be noted that the expansion of high speed railway networks is making rail increasingly attractive even for express services. The air

mode is both expensive and efficient in carrying relatively large loads over long distances. Due to the high costs, efforts are dedicated to the optimization of available capacity and the reduction of the number of aircrafts used. This translates into a relatively large number of alternatives in terms of aircraft types, capacities, and costs which adds to the complexity of the problem.

Kuby and Gray (1993) developed an early model for the design of the network of an express package delivery firm. It is a path-based $\{0,1\}$ network design model, where multi stop aircraft routes must be selected in and out of a given hub. Paths were generated *a priori* and the model was solved with a standard mixed-integer package. Analyses illustrated the cost effectiveness of a design with multiple stops over a pure hub-and-spoke network.

Kim *et al.* (1999) and Armacost, Barnhart, and Ware (2002) propose a more comprehensive model for the design of the multimodal version of the problem where both air and ground vehicles are considered (Barnhart and Schneur, 1996, address a simplified, single-hub version of the problem). In this problem, several hubs and aircraft types are considered, while trucks may perform pickup and delivery activities, as well as transportation over limited distances. The focus is on the design of the overnight air services for next-day delivery.

Two types of terminals are considered in the model: *gateways* where packages enter and exit the air network and hubs where packages are unloaded from in-coming airplanes, sorted, and loaded into an out-going aircraft for delivery to their destination gateways. Aircrafts may fly non-stop between gateways and hubs or stop at one gateway en-route. *Time window* restrictions on pickup and delivery times at gateways as well as on the sorting periods at hubs limit the number of stops and constrain the design. Aircrafts of different types are available in limited numbers. Each aircraft type has an operating cost and a capacity, as well as operating characteristics (type of engine, flying range, speed, etc.) that determine the routes it can fly. The objective is to design the minimum cost set of routes, aircraft assignments to these routes, and package flows to satisfy demand and level-of-service objectives, while complying with the operating parameters of the terminals (e.g., capacity), aircraft type (e.g., capacity, range, speed), and aircraft fleet (number of planes available).

A commodity $p \in \mathcal{P}$ represents packages to be moved from an origin gateway $o(p)$ to a destination gateway $d(p)$. Let d^p indicate the corresponding demand, $\mathcal{H}$ the set of hubs, and $\{o(p), d(p),\ p \in \mathcal{P}\}$ the set of origin and destination gateways. Let $\mathcal{F}$ be the set of aircraft types available and $\mathcal{R}^f$ the set of routes (sequence of gateways starting or ending at a hub) that may be flown by aircrafts of type $f \in \mathcal{F}$. The routes become the flight arcs $\mathcal{A}$ linking the hub and gateway nodes $\mathcal{N}$ of the system. The cost of flying route $r \in \mathcal{R}^f$ with aircrafts of type $f \in \mathcal{F}$ is denoted c_r^f, while u_r^f gives its capacity. Costs associated to package flows are not significant compared to those of operating the air fleets and are thus not included in the formulation. The number of available aircrafts of type f is n_f and the landing capacity of hub $h \in \mathcal{H}$ is a_h. Three indicators are also defined: 1) $\delta_{ij}^{rf} = 1$ when flight arc (i,j) is included in route r flown by aircraft type f and 0, otherwise; 2) $\beta_i^r = 1$ (-1) if node i is the origin (destination) gateway of route r and 0, otherwise; and 3) $\delta_r^h = 1$ if hub h is included in route r and 0, otherwise.

Two types of decision variables are defined: y_r^f indicates the number of times route $r \in \mathcal{R}^f$ is flown by aircrafts of type $f \in \mathcal{F}$; continuous variable x_{ij}^p stands for the amount of product p on the air link $(i,j) \in \mathcal{A}$. The basic frequency service network design formulation is then:

$$\text{Minimize} \quad \sum_{f \in \mathcal{F}} \sum_{r \in \mathcal{R}^f} c_r^f y_r^f \tag{37}$$

$$\text{subject to} \quad \sum_{j \in \mathcal{N}} x_{ij}^p - \sum_{j \in \mathcal{N}} x_{ji}^p = \begin{cases} d^p & \text{if } i = o(p), \\ -d^p & \text{if } i = d(p), \\ 0 & \text{otherwise,} \end{cases} \qquad i \in \mathcal{N},\ p \in \mathcal{P} \tag{38}$$

$$\sum_{p \in \mathcal{P}} x_{ij}^p \leq \sum_{f \in \mathcal{F}} \sum_{r \in \mathcal{R}^f} \delta_{ij}^{rf} u_r^f y_r^f \qquad (i,j) \in \mathcal{A} \tag{39}$$

$$\sum_{r \in \mathcal{R}^f} y_r^f \leq n_f \qquad f \in \mathcal{F} \tag{40}$$

$$\sum_{r \in \mathcal{R}^f} \beta_i^r y_r^f = 0 \qquad i \in \mathcal{N},\ f \in \mathcal{F} \tag{41}$$

$$\sum_{r \in \mathcal{R}^f} \delta_r^h y_r^f \leq a_h \qquad h \in \mathcal{H} \tag{42}$$

$$y_r^f \geq 0 \text{ and integer} \qquad r \in \mathcal{R}^f,\ f \in \mathcal{F} \tag{43}$$

$$x_{ij}^p \geq 0 \qquad (i,j) \in \mathcal{A},\ p \in \mathcal{P} \tag{44}$$

Constraints (38) enforce flow conservation for each product, while the forcing constraints (39) restrict the flow on each fly arc to the capacity of the aircrafts flown on that route. Constraints (40) and (42) enforce the number of available aircrafts of each type and the landing capacities at hubs, respectively. Constraints (41) are the aircraft balance restrictions: the number of aircrafts of each given type landing at a location (delivery routes) must equal the number taking off from that same location (pick up routes).

In Kim *et al.* (1999), the authors examine arc, path, and tree-based formulations of this model, and select the latter since it yields a problem of significantly reduced size compared to the two others. The authors solve the linear relaxation of the resulting formulation by combining heuristics, which further reduce the size of the problem, cut-set inequalities to strengthen the relaxation, and column generation to gradually generate a good set of route variables. Branch-and-bound is then used to obtain an integer solution. A different methodological approach is presented by Armacost, Barnhart, and Ware (2002). The authors transform the problem formulation by defining new composite variables that combine the original air service frequency and the package flow variables to represent possible air routes from gateway to hub with minimal but sufficient capacity to transport the required volume. The composite-variable formulation includes constraints that force the selection of at least one composite variable for each gateway-hub connection. The authors take advantage of the fact that air routes pass at most through two gateways and thus the number of composite variables, which allow to capture multiple aircraft routes with a single variable, is relatively limited. Moreover, the composite route variables implicitly account for the flow distribution

and thus yield a pure design formulation for which stronger bounds and thus more efficient solution methods may be derived. The methodology has been implemented at a major U.S. express package delivery company with significant success (Armacost *et al.*, 2004). The model is used on a continuous basis for the planning of next-day operations as well as for what-if scenario analyses. Such results emphasize the need to continue to explore the network design formulations for new insights and more efficient solution methods.

The reorganization of the German postal services belongs to the same problem class, albeit on a more comprehensive scale. Grünert and Sebastian (2000; see also Grünert, Sebastian, and Thärigen, 1999 and Buedenbender, Grünert, and Sebastian, 2000) decompose the problem into several subproblems of manageable proportions: the optimization of the overnight airmail network, the design of the ground-feeding and delivery transportation system, and the scheduling of operations. Vehicle routing models and techniques, as well as a discrete, time-dependent network design formulation (see Section 3.2.3), are proposed for the routing and scheduling tasks. The air network design formulation is further decomposed into a direct flight problem and a hub system problem; both subproblems are fixed cost, multicommodity, capacitated network design formulations with side constraints. To optimize these formulations, the authors propose combinations of classical heuristics, tabu search and evolutionary meta-heuristics, and exact mathematical programming methods (branch-and-bound). A decision support system integrates the models and associated solution methods, as well as the tools required to handle the data, models, and methods, and to assist in the decision process.

3.2.3 Time-Dependent Service Network Design

When schedules are contemplated, a *time* dimension must be explicitly introduced into the formulation. This is usually achieved by representing the operations of the system over a certain number of *time periods* by using a *space-time* network. The representation of the physical network is replicated in each period. Starting from its origin in a given period, a service arrives (and leaves, in the case of intermediary stops) at a later period at other terminals. Services thus generate temporal service links between different terminals at different time periods. Temporal links that connect two representations of the same terminal at two different time periods may represent the time required by terminal activities or the freight waiting for the next departure. The costs associated with the arcs of these networks are similar to those used in the static formulations of the previous subsections. Additional arcs may be used to capture penalties for arriving too early or too late.

There are again two types of decision variables. Integer design variables are associated with each service. Restricted to $\{0, 1\}$ values, these variables indicate whether or not the service leaves at the specified time. When several departures may take place in the same time period, general (nonnegative) integer variables must be used. (Note that one can always use $\{0, 1\}$ variables only by making the time periods appropriately small.) Continuous variables are used to represent the distribution of the freight flows through this service network.

The resulting formulations are network design models similar to those presented in the previous subsections but on significantly larger networks due to the time dimension. The sheer size of the space-time network, as well as the additional constraints usually required by the time dimension, makes this class of problems more difficult to solve than static versions. A limited number of contributions have thus been reported to date. Most address the problem in the context of a single carrier and propose heuristic solution methods (e.g., Farvolden and Powell, 1991, 1994; Farvolden, Powell, and Lustig, 1992; Equi *et al.*, 1997).

Different formulations are being proposed to address the recent trends in the development and operation policy of rail intermodal networks. One such trend is the introduction of booking systems that force customers to book in advance a precise number of container space slots on a given day and train service. To operate such systems, intermodal rail carriers enforce regular and cyclically scheduled services. Moreover, in order to decrease operating costs (by simplifying operations and reducing equipment handling) and to increase the utilization of their assets, rail cars and engines, mainly, the characteristics of services in terms of composition, capacity, and so on, are supposed to stay the same for all the time the schedule is valid. This starts to be known as a "full-asset-utilization" operating policy.

Not much work has yet been done with respect to these new operation characteristics, which may be observed both in North America and Europe. Pedersen, Crainic, and Madsen (2006) present one of the first contributions. The authors notice that a "full-asset-utilization" operating policy requires that the asset circulation issue be integrated into the service network design model. They represent this requirement by enforcing the condition that at each node of the representation, the (integer) design flow be balanced. The authors propose two formulations (a link and a cycle-based one) and a tabu search-based meta-heuristic that appears to efficiently yield good quality solutions. Much more work is required in this area, however.

3.3 Port Dimensioning

Important strategic planning issues for the design of container port terminals are related to the number of berths, the size of storage space, and the number of pieces of various equipment to install. Port dimensioning issues require a trade-off between the amount of investment and the level of customer service. For example, as the number of berths increases, the turnaround (waiting) time of vessels decreases. Because of the high investment and operating cost of container ships, delays experienced at a port generate high costs to the ship's operators and cause delayed arrivals at successive ports creating serious downstream operational problems. Thus, availability of empty berths at a port is a key issue when operators determine container terminals for their ships. Studies aimed at determining the optimal number of berths to provide usually consider the costs related to the turnaround time of ships, the berth construction cost, and the berth operation cost.

Quay cranes are the most expensive (5 to 10 million dollars per quay crane) handling

equipment in port container terminals. Determining the number of quay cranes to be installed is thus another major container-port design issue. In addressing this issue, one must consider that as the number of quay cranes per berth increases, the throughput rate per berth increases as well, but at a declining rate due to interference between adjacent quay cranes. Determining the number of pieces of other types of equipment, such as transporters, yard cranes, and so on, are also important port dimensioning issues.

Storage space is another critical resource in port container terminals, especially in major Asian hub ports such as Singapore, Hong Kong, Busan, Kobe, and Kaohsiung. As the storage space becomes smaller, the storage stacks become higher, which results in lower productivity of the transfer operations in the yard. Thus, there is a trade-off relationship between the investment in storage space and the productivity of the transfer operations in the yard (see Section 5).

Very few optimization-based models have been proposed for these issues, however. Descriptive models, mainly based on queuing theory and probabilistic approaches, have been used in most cases. The goal was to evaluate the performance of particular components of container terminals undervarious hypotheses regarding the number, type and, eventually, combination of resources, e.g., Fratar, Goodman, and Bran (1960), Miller (1971), Wanhill (1974), and Daganzo (1990) for the number of berths; Griffiths (1976) and Daganzo (1989a) for the number of quay cranes; Schonfeld and Sharafeldien (1985) for the number of quay cranes and berths; van Hee and Wijbrands (1988), Kozan (1997), and Kim and Kim (2002) for combinations of different types of resources including yard cranes, storage slots, and yard trailers. Two exceptions to this trend are nonlinear programming model proposed by Noritake and Kimura (1990) for berth dimentioning and the maximum flow network model introduced by Vis, de Koster, and Roodbergen (2001) to determine the minimum number of automated guided vehicles for completing a given set of delivery tasks.

Simulation models have also been proposed to evaluate seaport container terminal design (e.g., number and layout of various resources), operation policies (e.g., working hours), operation characteristics (e.g., travel and handling times of various equipments) and general performance: Ramani (1996), Lai and Leung (2000), Nam, Kwak, and Yu (2002), etc. A different simulation approach was proposed by Alessandri, Sacone, and Siri (2006), which represent containers and their movements in a terminal as a set of queues. Discrete-time equations are then used to describe the dynamic evolution of the system, where control variables represent the utilization of the terminal resources. This model yields an optimization problem that aims to minimize the transfer delays in the terminal. The problem is stated as an optimal control problem and is addressed by a receding-horizon solution strategy.

4 Container Fleet Management

The need for freight carriers to move empty vehicles follows from the differences in demand and supply for each commodity observed at most locations, resulting in an accumulation of empty vehicles in regions where they are not needed and in deficits of vehicles in other regions that require them. Vehicles must then be moved empty, or additional loads must be found, in order to bring them where they are needed to satisfy known and forecast demand in the following planning periods. This operation is known as *repositioning* or *empty balancing* and is a major component of what is known as *fleet management*. In its most general form, fleet management covers the whole range of planning and management issues from procurement of power units and vehicles to vehicle dispatch and scheduling of crews and maintenance operations. Often, however, the term designates a somewhat restricted set of activities: allocation of vehicles to customer requests and repositioning of empty vehicles. We follow this definition is this section.

Fleet management belongs to what is usually called the operational level of planning. Most system elements, demand, travel and handling times, and so on, vary with time. Most strategic and tactical planning models do not explicitly account for these variations. At the operational planning level, however, the time-dependency of data,decisions, and operations must be explicitly addressed, including the representation of the outcome of current decisions on the future state of the system. The *time-dependent* (*dynamic*) aspect of operations is further compounded by the *stochasticity* inherent to most systems, that is, by the set of uncertainties that are characteristic of real-life operations and management activities. If traffic is slower than predicted, vehicles may arrive late at customers' locations or at the terminal. Forecast customer requests for empty containers may not materialize while unexpected demands may have to be satisfied. The planned supplies of empty vehicles at depots may thus be unsettled and additional empty movements may have to be performed. Increasingly, these characteristics are reflected in the models and methods aimed at operational planning and management issues, as illustrated by the fleet management models of this section.

Moving vehicles empty does not directly contribute to the profit of the firm but it is essential to its continuing operations. Consequently, one attempts to minimize empty movements within the limits imposed by the demand and service requirements. Early models were based on transportation formulations or, most often, deterministic time-dependent transshipment network models (e.g., White, 1972 and Ermol'ev, Krivets, and Petukhov, 1976 for container fleet management). Details on the early approaches may be found in Dejax and Crainic (1987). Models that explicitly consider uncertainties in empty vehicle distribution started to be proposed in the mid 1980s' (e.g., Jordan and Turnquist, 1983). There is now a significant body of work in this area, mostly dedicated to rail and motor carrier issues (see the reviews by Crainic, 2003; Cordeau, Toth, and Vigo, 1998; Powell, 2003, Powell and Topaloglu, 2003, 2005; Powell, Bouzaïene-Ayari, and Simaõ, 2006).

Few efforts were dedicated to container fleet management issues. Crainic, Gendreau, and Dejax (1993) proposed a series of models for the allocation and management of a heteroge-

neous fleet of containers where loaded movements are exogenously accepted. Cheung and Chen (1998) focused on the single-commodity container allocation problem for liner regular ocean navigation lines) operators. Powell and Carvalho (1998) addressed the problem of the combined optimization of containers and flatcars for rail intermodal operations.

The container fleet management problem addressed by Crainic, Gendreau, and Dejax (1993) was cast in the context of the land transport part, as described in Section 3.1.1. The land container distribution system is composed of a number of port terminals, in-land depots, and customer locations. Ships arrive in ports carrying loaded and empty containers of various types and dimensions. Loaded containers are delivered to their destinations, using rail and truck under various types of contracts, while empty containers are available for delivery to customers in the vicinity of the port or for repositioning. Customers receiving loaded containers unload them and signal that they may be picked up and transported to a designated terminal (port or in-land). Similarly, customers that require empty containers of specific types for future shipments receive them from an in-land or a port terminal. If the type and number of containers requested by the customer are not available for delivery within the time window specified by the customer, the company can lease (or borrow from partner companies) containers or substitute certain types of containers for one another. The empty containers available for distribution to customers thus come from four sources: customers that unloaded their goods, world-wide repositioning, leasing, and substitution. Within the systems, containers are repositioned through two mechanisms: 1) the balancing movements determined by the strategic/tactical planning of the operations (Section 3.1.1), and 2) the allocation decisions concerning the depots to which the containers that become empty at customer locations are to be sent. The problem is dynamic, in the sense that demands vary in time and transport and customer operations may both require more than one period (day). It is stochastic due to variations in customer demands, including the possibility of demands from "unknown" customers, variations in the time required by customers to unload and return containers, and damage to containers (travel time variations were not considered in the paper).

The formulation is defined for a planning horizon discretized in T periods: $t = 1, 2, ..., T$. Let $\mathcal{H}$ be the set of ports and $\mathcal{D}$ the set of inland depots. The sources of randomness considered in the formulation are

d_i^p: Demand of containers of type p for demand customer $i \in \mathcal{I}^p$; The containers must be delivered within a time window Δ_i;

d_h^{pt}: Demand of empty containers of type p at port $h \in \mathcal{H}$ in period t for export;

o_s^{pt}: Supply of empty containers of type p released by supply customer s at time t; The set of these customers is denoted $\mathcal{S}^{pt}$;

o_h^{pt}: Supply of empty containers of type p arriving at port h in period t.

The parameters of the problem are

$\mathcal{J}_i^{pt}$: Set of depots j that may send a shipment of containers of type p that would arrive in period t at customer i;

$\mathcal{I}_j^{pt}$: Set of customer requests that may be served (i.e., he containers of type p will arrive within Δ_i) by a shipment starting from depot $j \in \mathcal{D} \cup \mathcal{H}$ in period t;

c_{ji}^{pt}: Unit transportation cost for a container of type p from a depot j to demand customer $i \in \mathcal{I}_j^{pt}$;

$\mathcal{S}_j^{pt}$: Set of supply customers from where containers of type p may reach terminal j in period t;

c_{sj}^{pt}: Unit transportation cost for a container of type p from supply customer $s \in \mathcal{S}_j^{pt}$ to a depot j;

c_{jk}^{pt}: Unit transportation cost for a container of type p between depots $j, k \in \mathcal{D} \cup \mathcal{H}$;

c_{pr}: 1/(Number of containers of type p needed to replace one container of type r);

c_j^{prt}: Unit cost at depot j in period t to substitute a container of type p for a container of type r;

c_j^{pt}: Unit holding cost for a container of type p at depot j in period t;

$\bar{c}_j^{pt}$: Cost of leasing (or borrowing) a container of type p at depot j in period t; It also represents the penalty cost at port h when d_h^{pt} cannot be satisfied.

With decision variables

x_{ji}^{pt}: Number of equivalent containers of type p allocated in period t from depot j to customer i;

x_j^{pt}: Number of containers of type p allocated as type p in period t from depot j;

x_j^{prt}: Number of containers of type p substituted for containers of type r in period t at depot j;

x_{sj}^{pt}: Number of containers of type p picked up at customer s in period t and delivered to depot j in period $\bar{t} \geq t$;

w_j^{pt}: Number of containers of type p in stock at depot j at the end of period t;

w_{jk}^{pt}: Number of containers of type p repositioned from depot j to depot k in period t;

b_j^{pt}: Number of containers of type p rented at depot j in period t;

the model minimizes the expected total operating cost, including substitutions and stockouts, over a multi-period planning horizon

$$E\left[\sum_{t}\sum_{p\in\mathcal{P}}\sum_{j\in\mathcal{H}\cup\mathcal{D}}\left(\sum_{s\in\mathcal{S}_j^{pt}} c_{sj}^{pt}x_{sj}^{pt}+\sum_{i\in\mathcal{I}_j^{pt}} c_{ji}^{pt}x_{ji}^{pt}+\sum_{k\in\mathcal{H}\cup\mathcal{D}} c_{jk}^{pt}w_{jk}^{pt}+\sum_{r\in\mathcal{P}} c_j^{prt}x_j^{prt}+c_j^{pt}w_j^{pt}+\bar{c}_j^{pt}b_j^{pt}\right)\right] \tag{45}$$

subject to non-negativity constraints on all variables, bounds on balancing flows, and constraints representing the dynamics of the system:

$$\sum_{t\in\Delta_i}\sum_{j\in\mathcal{J}_i^{pt}} x_{ji}^{pt}-d_i^p=0 \quad i\in\mathcal{I}^p \tag{46}$$

$$\sum_{\bar{t}\geq t}\sum_{j\in\mathcal{S}_{jpt}} x_{sj}^{pt}-o_s^{pt}=0 \quad s\in\mathcal{S}^{pt},\ p\in\mathcal{P},\ t=1,2,....,T \tag{47}$$

$$x_j^{pt}+\sum_{r\in\mathcal{P}} c_{pr}x_j^{prt}-\sum_{i\in\mathcal{I}_j^{pt}} x_{ji}^{pt}=0 \quad j\in\mathcal{D}\cup\mathcal{H},\ p\in\mathcal{P},\ t=1,2,...,T \tag{48}$$

$$w_j^{pt}+x_j^{pt}+x_j^{prt}+\sum_{k\in\mathcal{H}\cup\mathcal{D}} w_{jk}^{pt}-w_j^{pt-1}-\sum_{s\in\mathcal{S}_j^{pt}} x_{sj}^{pt}-\sum_{\bar{t}\leq t}\sum_{k\in\mathcal{H}\cup\mathcal{D}} w_{kj}^{\bar{t}t}-b_j^{pt}=0$$
$$j\in\mathcal{D},\ p\in\mathcal{P},\ t=1,2,....,T \tag{49}$$

$$w_h^{pt}+x_h^{pt}+x_h^{prt}+\sum_{k\in\mathcal{H}\cup\mathcal{D}} w_{hk}^{pt}+d_h^{pt}-w_{hk}^{pt}-\sum_{s\in\mathcal{S}_h^{pt}} x_{sh}^{pt}-\sum_{\bar{t}\leq t}\sum_{k\in\mathcal{H}\cup\mathcal{D}} w_{hk}^{pt}-o_h^{pt}-b_h^{pt}=0$$
$$h\in\mathcal{H},\ p\in\mathcal{P},\ t=1,2,...,T \tag{50}$$

where constraints (46) and (47) enforce demand and supply conditions at customers, equations (4) and (4) are the flow conservation conditions at in-land depots and ports, respectively, while equations (48) represent the allocation accounting between true and substitution containers at depots. The authors presented single and multicommodity deterministic formulations and a two-stage, restricted-recourse single commodity, stochastic model.

The stochastic and dynamic empty container allocation model of Cheung and Chen (1998) addresses the problem from the point of view of a container liner company that offers regular service lines between a given number of ports. Ships transport loaded containers that bring a profit and, space permitting, empty containers to reposition them in order to be able to satisfy future forecast demand. It is assumed that only one ship travels between two ports at each time period and that demand for loaded containers between ports does not exceed the capacity of the ship. Only one product - empty containers - is considered. The schedule of ships linking the ports in $\mathcal{H}$ is assumed fixed for the planning horizon $t=1,2,...,T$. Similarly to the formulations of Crainic, Gendreau, and Dejax (1993), one supposes that the

model is to be run into a rolling horizon mode: the model gives a solution (that accounts for the impact of today's decisions on the future state of the system), the suggestion solution for the first period is implemented and then, the next period, the model is run again using the up-dated information.

Three sources of randomness are considered in the model:

κ_{ij}^t: residual capacity for empty containers on the ship traveling from port $i \in \mathcal{H}$ to port $j \in \mathcal{H}$, leaving at time t;

δ_i^t: Demand for containers at port i at period t;

σ_i^t: Supply of containers at port i at period t (before unloading from ships).

The parameters of the problem are

c_{ij}^t: Unit transportation cost from port $i \in \mathcal{H}$ to port $j \in \mathcal{H}$, leaving at time t;

l_i^t: Unit cost for loading a container on a ship at port i at period t;

u_i^t: Unit cost for unloading a container from on a ship at port i at period t;

c_i^t: Unit holding cost for a container at port i in period t;

$\hat{c}_i^t$: Unit cost of leasing (or borrowing) at port i in period t;

R_i^t: Unit revenue from satisfying demand at port i at period t;

τ_{ij}: Transportation time from port i to port j.

and the decision variables

$x_i^{l,t}$: Number of containers loaded at port i in period t for any destination;

$x_i^{u,t}$: Number of containers unloaded at port i in period t from any other port;

$x_{ij}^{c,t}$: Number of containers to be repositioned from port i to port j departing in period t;

$x_{ij}^{p,t}$: Number of containers currently on the ship, in port i at period t, with destination port j;

$x_i^{h,t}$: Number of containers stored at port i at period t;

$x_i^{r,t}$: Number of leased containers at port i and period t used to meet demand;

$x_i^{d,t}$: Number of owned containers at port i and period t used to meet demand.

All decision variables are non-negative. The objective is to minimize the expected total cost while maximizing the revenue from satisfying demand:

$$\text{Minimize } E\left[\sum_{t}\sum_{i\in\mathcal{H}}\left(l_i^t x_i^{l,t} + u_i^t x_i^{u,t} + c_i^t x_i^{h,t} + \bar{c}_i^t x_i^{r,t} + \sum_{j\in\mathcal{H}} c_{ij}^t x_{ij}^{c,t} - R_i^t \delta_i^t\right)\right] \tag{51}$$

while the following relations define the system dynamics and constraints of the model:

$$\sum_{k\in\mathcal{H}} x_{ki}^{c,t-\tau_{ij}} - x_i^{u,t} - \sum_{j\in\mathcal{H}} x_{ij}^{p,t} = 0 \qquad i\in\mathcal{H},\ t=1,2,...,T \tag{52}$$

$$x_i^{h,t} + \delta_i^t - x_i^{h,t-1} - \sigma_i^t - x_i^{h,t} = 0 \qquad i\in\mathcal{H},\ t=1,2,...,T \tag{53}$$

$$\sum_{j\in\mathcal{H}} x_{ij}^{c,t} - x_i^{l,t} - \sum_{j\in\mathcal{H}} x_{ij}^{p,t} = 0 \qquad i\in\mathcal{H},\ t=1,2,...,T \tag{54}$$

$$x_{ij}^{c,t} \le \kappa_{ij}^t \qquad i\in\mathcal{H},\ t=1,2,...,T \tag{55}$$

$$x_i^{r,t} = \delta_i^t - x_i^{d,t} \qquad i\in\mathcal{H},\ t=1,2,...,T. \tag{56}$$

Relations (55) specify that demand must be satisfied. At each port and period, equations (52) compute the volume of containers unloaded from ships (the difference between the total volumes arriving and departing on all ships), relations (53) represent the inventory equations that also enforce the requirement that demand must be satisfied, and equations (54) compute the number of containers being repositioned. Constraints (55) enforce the residual capacity for repositioning movements and equations (56) defines the number of leased containers as the difference between the number of owed containers available and demand. The formulation is represented by a network problem with random arc capacities and is cast as a two-stage stochastic program. Stochastic quasi-gradient and hybrid approximation solution procedures are proposed and are compared through a rolling-horizon experiment.

Powell and Carvalho (1998a) address fleet management problems within the context of rail intermodal systems. The goal is to consider simultaneously the management of the fleet of flat cars the railway uses to provide service, the fleet of trailers and containers (collectively know as boxes) the railway owns and rents to its customers, and the complex rules that govern the substitutions among box types and the assignment of boxes to flat cars. The problem is highly dynamic and characterized by two sources of stochasticity: forecast of customer demand and the return of boxes and flat cars from customers and other railways.

The authors decompose the problem into a problem addressing the management of the boxes the railway owns and a flat car fleet management problem. Each is formulated as a mixed-integer program over a multicommodity space-time diagram representing "all" possible movements of vehicles and loads (including holding activities) over the planning horizon. Decision variables concern departure times, the type of vehicle (and departure time) for each load, repositioning movements, flows of vehicles and loads. The objective maximizes the expected total profit of the system. Each formulation is then cast as a recursive dynamic model and, using approximations of the future values of vehicles and boxes at the nodes of

the space-time network, it is decomposed into "easy-to-solve" local subproblems (assignment of boxes to requests or combinations of boxes to flat cars). In a series of forward-backward passes through the network, the algorithm then refines these approximations and assignments. Experimentations based on actual data suggest significant improvements over the planning procedures used in industry. More details on this methodology may be found in Powell (2003), Powell and Topaloglu (2003, 2005), and Powell, Bouzaïene-Ayari, and Simaõ (2006).

5 Models for Seaport Container Terminal Operations

This section describes issues and introduces corresponding models for operational planning and control in port container terminals. The following subsections introduce models for berth scheduling, quay-crane scheduling, stowage planning and sequencing, storage space planning, and dispatching of yard cranes and transporters: yard trucks, straddle carriers and, for automated terminals, automated guided vehicles (AGV). Reviews can be found in Steenken, Voß, and Stahlbock (2004) and in the book edited by Günther and Kim (2005).

5.1 Berth Scheduling

As already mentioned, berths are the most important resource in port container terminals because berth construction costs are the highest among all relevant cost factors. *Berth scheduling* is the process of determining the time and position at which each arriving vessel will berth. *Quay-crane allocation* is the process of determining the vessel that each quay crane will serve and the time during which the quay crane will serve the assigned vessel. (The terms *scheduling* and *allocation* are often used interchangeably for both problems.) The berth scheduling and the quay-crane allocation problems are related because the number of quay cranes assigned to a vessel impacts the berthing duration of the ship. Despite this relationship, most studies treat the two issues separately to avoid the complexity of the integrated problem. The study by Park and Kim (2003) is an exception.

Ports have long used priority rules to determine the allocation of berths to incoming ships and the earliest studies focused on this approach (e.g., van der Heyden and Ottjes, 1985). Simulation was used to evaluate and compare rules (e.g., Lai and Shih, 1992). Brown, Lawphongpanich, and Thurman (1994) and Brown *et al.* (1997) proposed the first mathematical models for allocating berths to vessels. Their studies focused on military vessels and assumed that each vessel required different services (re-provisioning, maintenance, repair, training, and certification test). Since not all services were provided at all berths, scheduling vessel shifts between berths was an important issue. This is different from the case of container ships for which only loading and unloading operations need to be considered.

Many ports are configured such that berths may be considered as subsections of a continuous line that ships of finite lengths can share, and several studies (Lim, 1998; Park and Kim, 2002; Park and Kim, 2003; Kim and Moon, 2003; and Guan and Cheung, 2004) consider the berth allocation problem as a type of continuous line partitioning problem. Most researchers have treated berth scheduling as a discrete resource allocation problem, however. This approximation reduces the problem to that of assigning berths to arriving ships, which is much easier to address than the continuous berth-scheduling problem. We start the presentation with this second approach.

Cordeau *et al.* (2003 proposed an integer programming model for the discrete version of

the berth-allocation problem based on a Multi-Depot Vehicle Routing Problem with Time Windows (MDVRPTW) formulation, where ships are seen as customers and berths as depots. One vehicle is located at each depot. Each vehicle starts and ends its tour at its depot. Ships are modeled as vertices in a multi-graph. Every depot is divided into an *origin* vertex and a *destination* vertex. Time windows can be imposed on every vertex to represent the availability period of the corresponding berth. Notation is as follows:

N: Set of ships; $n = |N|$;

M: Set of berths; $m = |M|$;

t_i^k: Handling time of ship i at berth k;

a_i: Arrival time of ship i;

s^k: Start of availability time of berth k;

e^k: End of availability time of berth k;

b_i: Upper bound of the service time window of ship i;

v_i: Value of the service time for ship i;

$o(k)$: Starting operation time of berth k;

$d(k)$: Ending operation time of berth k;

$M_{ij}^k = \max\{b_i + t_i^k - a_j, 0\}, k \in M$, i and $j \in N$;

$G^k = (V^k, A^k), k \in M$, where $V^k = N \cup \{o(k), d(k)\}$ and $A^k \subseteq V^k \times V^k$.

With decision variables

x_{ij}^k: $x_{ij}^k = 1$, if and only if ship j is scheduled after ship i at berth k; 0, otherwise;

T_i^k: Berthing time of ship i at berth k, i.e., the time when the ship moors;

$T_{o(k)}^k$: Starting operation time at berth k, i.e., the time when the first ship moors at the berth;

$T_{d(k)}^k$: Ending operation time at berth k, i.e., the time when the last ship departs from the berth;

the MDVRPTW model may be written as:

$$\text{Minimize} \quad \sum_{i \in N} \sum_{k \in M} v_i \left[T_i^k - a_i + t_i^k \sum_{j \in N \cup \{d(k)\}} x_{ij}^k \right] \tag{57}$$

$$\text{subject to} \quad \sum_{k \in M} \sum_{j \in N \cup \{d(k)\}} x_{ij}^k = 1 \qquad i \in N \qquad (58)$$

$$\sum_{j \in N \cup \{d(k)\}} x_{o(k),j}^k = 1 \qquad k \in M \qquad (59)$$

$$\sum_{j \in N \cup \{o(k)\}} x_{i,d(k)}^k = 1 \qquad k \in M \qquad (60)$$

$$\sum_{j \in N \cup \{d(k)\}} x_{ij}^k - \sum_{j \in N \cup \{o(k)\}} x_{ji}^k = 0 \qquad k \in M, i \in N \qquad (61)$$

$$T_i^k + t_i^k - T_j^k \leq (1 - x_{ij}^k) M_{ij}^k \qquad k \in M, (i,j) \in A^k \qquad (62)$$

$$a_i \leq T_i^k \qquad k \in M, i \in N \qquad (63)$$

$$T_i^k + t_i^k \sum_{j \in N \cup \{d(k)\}} x_{ij}^k \leq b_i \qquad k \in M, i \in N \qquad (64)$$

$$s^k \leq T_{o(k)}^k \qquad k \in M \qquad (65)$$

$$T_{d(k)}^k \leq e^k \qquad k \in M \qquad (66)$$

$$x_{ij}^k \in \{0,1\} \qquad k \in M, (i,j) \in A^k \qquad (67)$$

The objective is to minimize the weighted sum of the service times for all the vessels. Constraints (58) implies that each vessel must be assigned once to a berth. Constraints (59) and (60) define the degree of the depots. Constraints (61) enforce the flow conservation. The consistency of the T_i^k variables with the time sequence on a berth is guaranteed by constraints (62). Constraints (63) and (64) indicate the service time windows for vessels, while constraints (65) and (66) set the available time windows on the berths. For small instances, the above formulation can be solved by commercial software for integer linear programming models. For problem instances of realistic size, Cordeau *et al.* (2003) proposed a tabu search-based meta-heuristic.

A different approach consists in explicitly representing and penalizing the difference between the berthing order implied by the ship priority and the one proposed. The resulting models take the form of non-linear integer programming formulations (Imai, Nagaiwa, and Tat, 1997 and Imai, Nishimura, and Papadimitriou, 2001, 2003). Lagrangian-based and, for problem instances of realistic size, genetic meta-heuristic solution methods have been proposed. The discrete berth-scheduling problem may also be cast as a machine-scheduling problem. Li, Cai, and Lee (1998) introduced a formulation for the scheduling of a single processor (the berth) that can simultaneously perform multiple jobs (vessels). The authors aimed to minimize the makespan and, based on the similarity of the problem to the bin-packing problem, suggested various algorithms based on the First-Fit-Decreasing heuristic. Guan *et al.* (2002) proposed an m-parallel machine scheduling formulation. In their model, the machines (quay cranes) are arranged along a straight line and each job (ship) requires simultaneous processing by multiple consecutive processors. Ships are characterized by size, processing time, and weight (priority). A heuristic was proposed to minimize the total weighted completion time of the jobs (ships).

A drawback of considering a berth as a collection of discrete berthing locations is that the number of ships that may be served simultaneously is fixed regardless of ship lengths. The continuous representation does not have this limitation: for the same length of berth, more vessels can be served simultaneously if they are shorter. On the other hand, in contrast to the discrete variant, the continuous berth-scheduling problem requires determining the exact berthing position of each ship as a real-valued position on a continuous line. Moreover, the berthing time for each ship must be determined simultaneously. The goals of the process include minimizing the ship departure delays and the container-handling costs that depend on the berthing position of each ship.

Lim (1998) was the first to consider a berth as a continuous line rather than a collection of discrete segments and viewed the berth planning problem as a two-dimensional bin packing problem. He discussed how to locate berthing positions of vessels so that the throughput of the berth is maximized, but did not consider the berthing time as a decision variable. Park and Kim (2002) introduced the first model for the continuous berth-scheduling problem to determine simultaneously the berthing time and position for each ship. The objective of the model is to minimize the costs resulting from delayed vessel departures, plus the additional handling costs resulting from deviations of the berthing position from the best location on the berth. The following information is assumed known:

L: Length of the berth;

l: Number of vessels;

a_i: Expected arrival time of vessel i;

b_i: The ship operation time required for vessel i;

d_i: Requested departure time of vessel i;

l_i: Length of vessel i. This value includes the required gap between adjacent vessels;

c_{1i}: Additional travel cost (per unit berth length) of delivering containers to vessel i resulting from non-optimal berthing locations;

c_{2i}: Penalty cost (per time unit) of vessel i resulting from a delayed departure after the requested due time.

The problem is set in a 2-dimensional space, the berth length and the planning time defining the horizontal and vertical axes, respectively. The reference point for the berth-length coordinate is the leftmost boundary of the berth. Figure 4 illustrates this setting, as well as the relationships between the input data and the following decision variables:

x_i: Berthing position of vessel i;

y_i: Berthing time of vessel i;

p_i: Best berthing location of vessel i; This location is represented by the x-coordinate (berth length axis) of the leftmost end of the vessel. It is determined by considering the distribution of containers to be loaded into the vessel;

z_{ij}^x: Equals 1, if vessel i is located to the left of vessel j in the time-berth-length space; 0, otherwise; Vessel i is located to the left of vessel j in Figure 4;

z_{ij}^y Equals 1, if vessel i is located below vessel j in the time-berth-length space; 0, otherwise; Vessel i is not considered to be located below vessel j in Figure 4 because of their partial overlap in the time dimension.

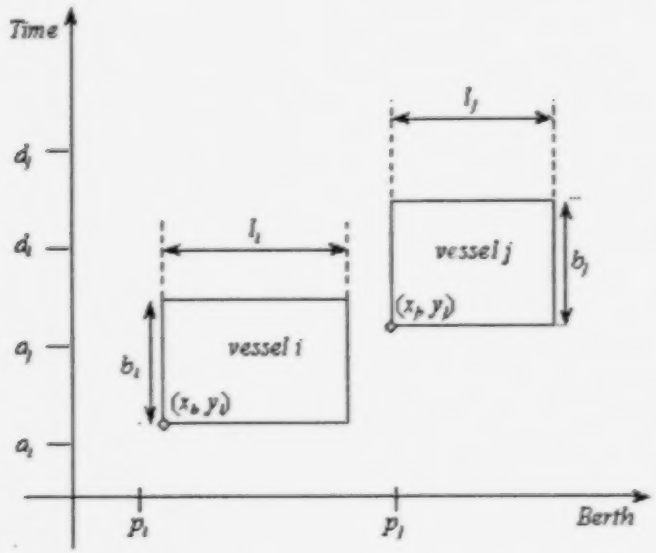

Figure 4: Structure of Continuous Berth Scheduling Problem (Park and Kim, 2002)

An objective function can then be written as:

$$\text{Minimize} \quad \sum_{i=1}^{I} \{c_{1i}|x_i - p_i| + c_{2i}(y_i + b_i - d_i)^+\}, \tag{68}$$

where $x^+ = \max\{0, x\}$. The first term of (68) computes the penalty for the deviation of the berthing positions from the best locations, while the second computes the penalty corresponding to the departure delays of the vessels leaving after the requested departure time. Let $|x_i - p_i|$ be α_i^+ when $x_i - p_i \geq 0$ and α_i^- when $x_i - p_i < 0$. Let $(y_i + b_i - d_i)$ be β_i^+ when $y_i + b_i - d_i \geq 0$ and β_i^-, otherwise. Then, the berth scheduling problem can be formulated as follows:

$$\text{Minimize} \quad \sum_{i=1}^{I} \left\{c_{1i}(\alpha_i^+ + \alpha_i^-) + c_{2i}\beta_i^+\right\} \tag{69}$$

$$
\begin{aligned}
\text{Subject to } & x_i - p_i = \alpha_i^+ - \alpha_i^- && i = 1, 2, \ldots, l && (70)\\
& y_i + b_i - d_i = \beta_i^+ - \beta_i^- && i = 1, 2, \ldots, l && (71)\\
& x_i + l_i \leq L && i = 1, 2, \ldots, l && (72)\\
& x_i + l_i \leq x_j + M(1 - z_{ij}^x) && j = 1, 2, \ldots, l, \quad i \neq j && (73)\\
& y_i + b_i \leq y_j + M(1 - z_{ij}^y) && i, j = 1, 2, \ldots, l, \quad i \neq j && (74)\\
& z_{ij}^x + z_{ji}^x + z_{ij}^y + z_{ji}^y \geq 1 && i, j = 1, 2, \ldots, l, \quad i < j && (75)\\
& y_i \geq a_i && i = 1, 2, \ldots, l && (76)\\
& \alpha_i^+,\ \alpha_i^-,\ \beta_i^+,\ \beta_i^-,\ x_i \geq 0 && i = 1, 2, \ldots, l && (77)\\
& z_{ij}^x,\ z_{ij}^y \in \{0, 1\} && i, j = 1, 2, \ldots, l, \quad i \neq j && (78)
\end{aligned}
$$

Constraints (70) and (71) define $\alpha_i^+, \alpha_i^-, \beta_i^+$, and β_i^-. Constraints (72) specify that the position of the rightmost end of vessel i is restricted by the length of the berth. A constraint of the type (73) or (74) is effective only when the corresponding z_{ij}^x or z_{ij}^y, respectively, equals one. The constraints define the relative "left" and "bellow" position of vessel i relative to that of vessel j in the time-berth-length space. Constraints (75) enforce the condition that two vessels cannot occupy the same space during the same time. The constraints proceed by excluding in the time-berth-length diagram the case $z_{ij}^x + z_{ji}^x + z_{ij}^y + z_{ji}^y = 0$, which corresponds to schedule overlapping for vessels i and j. Constraints (76) imply that vessels cannot berth before they arrive.

Model (69)-(78) is a mixed-integer programming formulation. Park and Kim proposed a sub-gradient optimization technique and solved problems with 13 to 20 ships in a few minutes of computational time on a Pentium II (233 MHz and 128 Mb RAM) computer. Similar formulations have also been proposed by Kim and Moon (2003) and Guan and Cheung (2004). Kim and Moon examined a stability property that berthing locations of vessels must satisfy to form an optimal solution and used it in a simulated annealing procedure. Guan and Cheung used a different objective function. They minimized the total weighted flow time instead of the penalties for delayed service and the deviation of berthing positions from the best locations of Park and Kim (2002). Guan and Cheung proposed a tabu search meta-heuristic based on pair-wise exchanges, which appears to perform well.

5.2 Quay-Crane Scheduling

Quay cranes are another important resource in container terminals. In practice, the quay-crane schedule is usually constructed by operation planners as part of ship plans. Two types of quay-crane scheduling problems have been defined according to the detail of the representation. The first problem simultaneously schedules the berth and the quay cranes by specifying the starting and ending times of unloading and loading operations for each quay crane assigned to a specific ship. Park and Kim (2003) proposed a model and a two-phase solution procedure. The first phase determines the berthing position and time of each

vessel as well as the number of quay cranes assigned to each vessel at each time segment. Sub-gradient optimization is used to obtain a near-optimal solution of the first phase. In the second phase, a detailed schedule for each quay crane is constructed by dynamic programming starting from the solution of the first phase.

QUAY CRANE SCHEDULE

QUAY CRANE NO. : 101
SHIP NAME : COLOMBO BAY
VOY NO : VLBY - 05

DATE : 12-12-2004
TIME : 11.40-18.00

WORKING ORDER	BAY	WORK	DISCHARGE			LOAD			SCHEDULE	
			20FT	40FT	45FT	20FT	40FT	45FT	FROM	TO
1	01 Hold	Loading				8			11 : 40	12 : 00
2	03 Deck	Loading				1			12 : 00	12 : 10
3	19 Hold	Loading				32			12 : 10	13 : 20
4	22 Hold	Loading					42		13 : 20	15 : 00
5	25 Hold	Loading				8			15 : 00	15 : 30
6	26 Deck	Loading					8		15 : 30	16 : 00
7	28 Hold	Loading				5			16 : 00	16 : 20
8	31 Hold	Loading				5			16 : 20	17 : 00
9	30 Hold	Loading					19		17 : 00	18 : 00

Figure 5: Example of Quay Crane Work Schedule

The second type of quay-crane scheduling problem determines the detailed schedule for each quay crane assigned to a vessel for a pre-specified time window to carry out container unloading or loading operations. Figure 5 illustrates such a schedule. Most studies on quay-crane scheduling address this problem. The modeling methodology generally used is based on integer programming models. Daganzo (1989b) addressed the quay-crane scheduling problem in port terminals for general and containerized cargo. He attempted to determine the number of quay cranes to assign to each ship bay at each time segment. An integer programming model was suggested with an objective to minimize the completion time of ship operations for all the ships.

Peterkofsky and Daganzo (1990) refined the previous model and proposed a branch-and-bound algorithm. In their model, S ships are moored alongside the berth of a terminal. Ship i has H_i holds numbered $(i,1),(i,2),\ldots,(i,H_i)$. The time required to complete the loading and unloading operations for hold (i,j) is denoted W_{ij} measured in quay-crane-hours. It is assumed that all ships are ready for loading/unloading operations at time $t=0$. It is also assumed that sufficient berth space is provided for all the ships to berth at the same time. M quay cranes are available. Decision variables are $y_{ijm}(t)=1$ if quay crane m, $m=1,2,\ldots,M$, is allocated to hold (i,j) at time t, and 0 otherwise. The objective

function minimizes the weighted amount of time that ships spend in port, $\sum_{i=1}^{S} C_i T_i$, where T_i and C_i stand for the departure time and cost per hour of delay for ship i, respectively. A ship can depart only when the operations on all its holds have been completed, that is, if

$$\sum_{m=1}^{M} \int_0^{T_i} y_{ijm}(t)dt \geq W_{ij} \quad j = 1, 2, \ldots, H_i. \tag{79}$$

In some cases, there may exist other activities than ship unloading and loading which must be performed (e.g., refill of supplies or repairs) and thus the earliest possible departure time may be restricted by a limit τ_i yielding the constraints

$$T_i \geq \tau_i \quad \text{for } i = 1, 2, \ldots, S. \tag{80}$$

A quay crane may be allocated to no more than one hold at a time. Thus, for any value of t, the crane assignments must satisfy:

$$\sum_{i=1}^{S} \sum_{j=1}^{H_i} y_{ijm}(t) \leq 1 \quad m = 1, 2, \ldots, M. \tag{81}$$

Finally, when only a maximum of Δ_{ij} quay cranes can be allocated to hold (i, j), the following constraints must be added to the formulation:

$$\sum_{m=1}^{M} y_{ijm}(t) \leq \Delta_{ij} \quad i = 1, 2, \ldots, S, \; j = 1, 2, \ldots, H_i. \tag{82}$$

Peterkofsky and Daganzo (1990) assumed a situation with no serious interference between cranes and, consequently, did not impose any restrictions on the movements of cranes. This approach is appropriate for general cargo handling. However, quay cranes in container terminals travel on the same rail, which results in various interference possibilities between adjacent cranes. Consequently, adjacent quay cranes need space of at least two ship bays between them. At the same time, no crane can pass over adjacent cranes, which makes the problem more complicated. Kim and Park (2004) addressed the quay-crane scheduling problem at this level of detail, but for a single vessel only. The model determines starting and ending times for each quay crane to serve each ship bay, under various constraints representing detailed movements of quay cranes and interferences among quay cranes. The authors proposed both a branch-and-bound algorithm and a GRASP-based heuristic to overcome the computational difficulty of solving the mixed-integer programming formulation exactly.

Comparing the work of Peterkofsky and Daganzo (1990) and Kim and Park (2004), one realizes that the difference comes from the different viewpoints on how the whole ship planning problem should be addressed. Peterkofsky and Daganzo (1990) considered the quay-crane scheduling as a part of berth planning, whereas Kim and Park (2004) separated the detailed crane-scheduling problem from the berth-scheduling problem. This comparison also provides an illustration of the often-encountered trade off between the integration of several planning issues into the same model and the level of detail one can include in the representation of each element of the system being represented.

5.3 Stowage Planning and Sequencing

A *container group* is defined as a collection of containers of the same size and with the same destination port. *Stowage planning* determines the block (cluster of adjacent slots) of a ship bay a specific group of containers should be stacked into. The stowage planning process must consider the burden of additional container manipulations when containers bound for succeeding ports are located in higher tiers than those that must be unloaded at a given port. The stowage plan must also comply with various measures of ship stability and strength. Figure 6 illustrates a stowage plan. Christiansen *et al.* (2006) review studies on stowage planning.

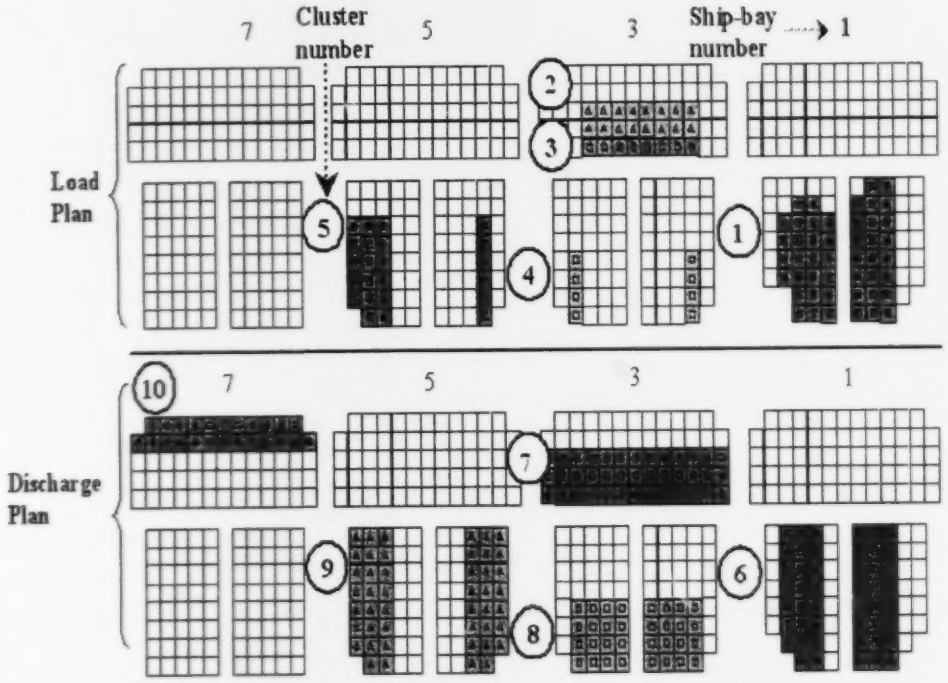

Figure 6: Stowage Plan Example (Partial View)

The *stowage-sequencing* problem must be addressed next by determining the sequence of unloading inbound containers and loading outbound containers. The slot into which each of the outbound containers will be stacked must be determined simultaneously. Figure 7 illustrates a loading sequence list. When an indirect transfer system is used, yard cranes move containers between yard stacks and yard trucks. The loading sequence of individual containers thus impacts significantly the total distance traveled by yard cranes and hence the total handling cost in the yard. This is not the case when a direct transfer system is used since the yard handling cost is determined by the cost of yard truck movements only, which is independent of the container loading sequence.

In the stowage (load/unload) sequencing problem, it is usually assumed that the stowage

LOADING SEQUENCE

QUAY CRANE NO. : 101
SHIP NAME : COLOMBO BAY
BAY NO. : 22 - HOLD

VOYAGE NO : CLBY-05

Sequence	Container No.	Yard Location	Size(Feet)	Full/Empty	Ship Location
001	TEXU5043257	4A-11-3-3	40	F	22-04-02
002	OCLU1504226	4A-11-3-2	40	F	22-08-04
003	TOLU2249587	4A-11-3-1	40	F	22-08-06
004	POCU4204642	4A-22-1-3	40	F	22-08-08
005	TRIU5332309	4A-22-1-2	40	F	22-08-10
006	OCLU1095339	4A-22-4-3	40	F	22-08-12
007	TEXU4962260	4A-22-2-3	40	F	22-04-04
008	GSTU7611558	4A-22-2-2	40	F	22-04-06
009	GSTU7610079	4A-22-3-3	40	F	22-04-08
010	POCU7013188	4A-22-3-2	40	F	22-04-10
011	TEXU5038271	4A-22-4-2	40	F	22-04-12

Figure 7: Example of a Loading Sequence List

plan is already constructed and provided by the vessel carrier. The problem is then reduced to assigning slots to individual loading containers and sequencing unloading and loading operations. Most research has focused on the sequence of loading operations, which influences the handling costs more significantly than the sequence of unloading operations. Indeed, containers to be loaded into the slots of a ship must satisfy various constraints on the slots, which are pre-specified by the stowage planner. Furthermore, locations of outbound containers may be scattered over a wide area in a terminal yard. The time required for loading operations thus depends on the cycle time of both quay and yard-cranes. The cycle time of a quay crane depends on the loading sequence of slots in the vessel, while the cycle time of a yard crane is affected by the loading sequence of containers in the yard.

Research on load sequencing can be classified according to the scope of the problem. Several efforts have addressed the *pickup-scheduling* problem in which the travel route of each yard crane and the number of containers to be picked up at each yard bay on the route are determined (Kim and Kim, 1999a; Kim and Kim, 1999b,c; Kim and Kim, 2003; Narasimhan and Palekar, 2002; Ryu *et al.*, 2001). Other authors have attempted to determine the loading sequence of individual containers present in a yard into the slots of a vessel, a process that requires more detailed scheduling than the previous problem (Cojeen and Dyke, 1976; Kozan and Preston, 1999; Kim, Kang, and Ryu, 2004). In this subsection, we focus on the pickup-scheduling problem, with only a brief literature review dedicated to the individual-container-scheduling problem.

The pickup-scheduling problem for a piece of yard equipment (e.g., a yard crane) can be described as follows (Kim and Kim, 1999b; Narashimhan and Palekar, 2002): In a terminal yard, containers are classified into groups, each yard bay holding a number of containers of a number of particular groups. Table 2 illustrates a yard map, that is, the distribution of containers by container group over the yard bays. It is a simplified case where each yard bay contains containers of one group only: 14 containers of group A are in yard-bay 1, 14 containers of group B are in yard-bay 2, and so on. A load plan specifies a sequence of blocks of pickup operations, each block representing a subsequence of consecutive pickups of containers belonging to the same group. Table 3 shows a load plan composed of eight subsequences to pick up sequentially 15 containers of group A, 20 containers of group B, 24 of group A, and so on. A partial tour denotes a sequence of yard bays the yard crane visits to perform the operations for a given subsequence. The pickup-scheduling problem is to decide (1) the container locations (yard bays) to use for each subsequence; and (2) the partial tour for each subsequence, i.e., the visiting order of the yard bays assigned to each subsequence in (1). The objective is to minimize the total setup and travel time the yard crane requires to pick up all the containers in the load plan.

The following pickup-scheduling model follows Kim and Kim (1999b). It addresses the scheduling of a single yard crane. The problem parameters are

m: Number of subsequences (and partial tours) that constitute a complete tour for a yard crane;

Yard-bay	1	2	4	5	7	8	9	11	12	14	15
Container group	A	B	C	B	A	B	C	B	A	C	A
Number of containers	14	14	13	14	12	7	12	8	13	10	13

Table 2: Yard Map

Subsequence number	1	2	3	4	5	6	7	8
Container group	A	B	A	C	B	A	B	C
Number of containers	15	20	24	25	11	23	12	10

Table 3: Load Plan

n: Number of yard bays;

l: Number of container groups;

c_{hj}: Initial number of containers of group h stacked at yard bay j;

t: Subsequence number, $t = 1, 2, \ldots, m$;

r_t: Number of containers in subsequence t;

g_t: Container-group number of subsequence t;

$S(h)$: Set of subsequences for the container-group number h;

$B(h)$: Set of the yard bays where containers of group h are stored;

S: Initial location of the yard crane, noted $B(g_0) = \{S\}$;

T: Final location of the yard crane, noted $B(g_{m+1}) = \{T\}$;

d_{ij}: Travel distance of the yard crane between yard bays i and j;

e_{ij}: Equals 1, if yard bays i and j are distinct and 0, otherwise (i.e., $i = j$);

T_s: Setup time of the yard crane for each visit to a yard-bay;

T_d: Travel time of the yard crane per the distance of a yard-bay length.

The decision variables are

y_{ij}^t = 1, if the yard crane moves from yard bay i to yard bay j after completing the partial tour t; 0, otherwise; y_{Sj}^0 and y_{jT}^m denote the first and the final movements of the yard crane during the tour, respectively;

z_{ij}^t = 1, if the yard crane moves from yard bay i to yard bay j during the partial tour t; 0, otherwise;

$x_j^t =$ Number of containers picked up at yard bay j during the partial tour t.

The problem may then be formulated as follows:

$$\begin{aligned}\text{Minimize } & T_s\left(\sum_{t=0}^{m}\sum_{i\in B(g_t),j\in B(g_{t+1})} e_{ij}y_{ij}^t+\sum_{t=1}^{m}\sum_{i,j\in B(g_t)} z_{ij}^t\right)\\ &+T_d\left(\sum_{t=0}^{m}\sum_{i\in B(g_t),j\in B(g_{t+1})} d_{ij}y_{ij}^t+\sum_{t=1}^{m}\sum_{i,j\in B(g_t)} d_{ij}z_{ij}^t\right) \qquad (83)\\ &=\sum_{t=0}^{m}\sum_{i\in B(g_t),j\in B(g_{t+1})}(T_se_{ij}+T_dd_{ij})\,y_{ij}^t+\sum_{t=1}^{m}\sum_{i,j\in B(g_t)}(T_se_{ij}+T_dd_{ij})\,z_{ij}^t\end{aligned}$$

$$\text{subject to } \sum_{j\in B(g_1)} y_{Sj}^0 = 1 \qquad (84)$$

$$-\sum_{j\in B(g_m)} y_{jT}^m = -1 \qquad (85)$$

$$\sum_{j\in B(g_{t-1})} y_{ji}^{t-1}-\sum_{j\in B(g_{t+1})} y_{ij}^t+\sum_{k\in B(g_t)}(z_{ki}^t-z_{ik}^t)=0 \qquad i\in B(g_t),\ t=1,2,\ldots,m \qquad (86)$$

$$\sum_{i,j\in B(g_t)} z_{ij}^t \le |N|-1 \qquad N\subseteq B(g_t),\ t=1,2,\ldots,m \qquad (87)$$

$$x_j^t \le M\left(\sum_{k\in B(g_t)} z_{kj}^t+\sum_{i\in B(g_{t-1})} y_{ij}^{t-1}\right) \qquad j\in B(g_t),\ t=1,2,\ldots,m, \qquad (88)$$

$$\sum_{j\in B(g_t)} x_j^t = r_t \qquad t=1,2,\ldots,m \qquad (89)$$

$$\sum_{t\in S(h)} x_j^t \le c_{hj} \qquad h=1,2,\ldots,l,\ j\in B(h) \qquad (90)$$

$$y_{ij}^t\in\{0,1\} \qquad i\in B(g_t),\ j\in B(g_{t+1}),\ t=1,2,\ldots,m \qquad (91)$$

$$z_{ij}^t\in\{0,1\} \qquad i,j\in B(g_t),\ t=1,2,\ldots,m \qquad (92)$$

$$x_j^t\ge 0 \qquad j\in B(g_t),\ t=1,2,\ldots,m, \qquad (93)$$

where M is a sufficiently large number and $|N|$ denotes the cardinality of set N.

The objective function (83) minimizes the total time required by the yard crane to perform the load plan, which depends on the total number of setups and the total distance traveled. Because a setup occurs whenever a yard crane moves from one yard bay to another, only the inter-yard-bay movements are considered in the evaluation of the total number of setups. A feasible solution corresponds to a path from the source node to the terminal node of the network. Constraints (84) and (85) respectively represent conservation of flow at the source and terminal nodes, respectively, while constraints (86) enforce the flow conservation at the other nodes. Relations (87) are subtour elimination constraints that exclude isolated cycles (i.e., not connected to a path from the source to the destination node) from a partial tour. Constraints (88) imply that only when a yard crane visits a yard bay can it pick up containers there. Constraints (89) enforce the condition that the number of containers picked up in a partial tour must equal the number of containers requested by the load plan. Constraints (90) indicate that the total number of containers picked up during the entire tour must not be larger than the initial number of containers available at each yard bay.

To build this formulation, Kim and Kim (1999b) took advantage of the limited motion of yard cranes, which move on a straight rail line, to reduce the solution space. Kim and Kim (1999a) and Kim and Kim (1999c) generalized this model for the pickup-scheduling problem for a single straddle carrier and for multiple straddle carriers, respectively. Kim and Kim (2003) compared exact optimization, a beam search heuristic, and a genetic meta-heuristic for solving the model proposed by Kim and Kim (1999b). Based on objective-function value, the beam search outperformed the genetic meta-heuristic on problems with 30 yard bays for which exact optimization algorithms could not be used due to excessive computing times. Ryu *et al.* (2001) suggested an ant system meta-heuristic for the same formulation and compared it to a tabu search meta-heuristic on problem instances with 39 to 514 ships and 36 to 350 yard bays. On average, the ant colony meta-heuristic was 49 times slower than the tabu search but achieved a solution value 17% better.

Narasimhan and Palekar (2002) analyzed a generalized version of the scheduling problem proposed by Kim and Kim (1999b). The authors relaxed the assumption that the order of subsequences is fixed, which may yield a reduction in the total handling time of transfer cranes. They also proved a number of properties of the optimal solution to this version of the problem, in particular that 1) the yard bays visited by a transfer crane during a partial tour must be contiguous in the yard-bay map; 2) sets of yard bays for partial tours of the same container group coincide with partitions of the yard-bay map; and 3) the minimization of the total distance traveled by transfer cranes results in the minimization of the transfer-crane setup time. The authors then restated the problem as an odd-node-matching problem where edges are added at minimum cost such that every node of the graph has en even degree. They proved that this version of the problem is NP-Hard in the strong sense and proposed both a branch-and-bound and an enumerative heuristic to address it.

The problem of constructing a containership loading plan at the level of individual containers appears more complex than the pickup-scheduling one. On the one hand, the problem

is larger because the loading sequence of individual containers must be determined. On the other hand, the goals and limitations related to the operation of various pieces of equipment, quay and transfer cranes notably, and by the constraints imposed by the stowage plan must be accounted for (Kim, Kang, and Ryu, 2004). The objectives related to the operation of quay cranes are first to fill slots in the same hold, then stack containers onto the same tier on deck, and finally stack containers of weights included in the same weight group as specified in the stowage plan. The objectives related to the operation of transfer cranes are to minimize the travel time of transfer cranes, minimize the number of rehandles (the same container that is moved more than once to, for example, allow access to containers stacked underneath), and pick up containers in locations nearer to the transfer point earlier than those located farther from the transfer point. The constraints related to the operation of quay cranes are to follow precedence relationships among slots that follow from their own work schedules and the relative positions of slots in a ship bay, not violate the maximum allowed total weight of the stack on deck, not violate the maximum allowed height of the stack of a hold, and load the same type of containers as specified in the stowage plan. A constraint related to the operation of transfer cranes is to maintain sufficient distance between adjacent transfer cranes to avoid interferences.

The early contributions presented software based on heuristic rules and simple procedures (e.g., Cojeen and Dyke, 1976). Kim, Kang, and Ryu (2004) proposed an integer linear programming model for sequencing individual outbound containers. The cost-minimization objective accounted for the handling cost of both quay and transfer cranes, while constraints included restrictions on the maximum weight allowed on deck, maximum height of stacks in holds, and the stability of the vessels. The authors proposed a beam search heuristic for this formulation. Kozan and Preston (1999) focused on the scheduling of transfer operations where the actual sequencing issues are not considered. The authors proposed a mixed-integer formulation to determine the storage locations of containers in the yard, the yard equipment to move the containers, and the schedules of these movements, such that the total travel cost of the yard equipment is minimized. A genetic meta-heuristic was proposed as solution method.

5.4 Storage Activities in the Yard

The scope of most research dedicated to the operation of yards is to minimize the container handling efforts and utilize the storage space efficiently. In addressing these issues, one must consider the rather different characteristics of inbound (usually import) and outbound (export) container flows. Inbound containers are usually unloaded fast not to delay the ship departure, but are retrieved over a long period of time compared to the ship unloading time, in a somewhat random sequence due to uncertainties related to importing formalities and the operations of land transportation modes. The minimization of the number of container rehandlings is thus the most important issue in this case. The same characteristics of the land and maritime transportation modes make outbound containers arrive at the yard randomly over a long period of time, while their loading is performed rapidly. Pre-planning storage

spaces for arriving containers will result in less rehandlings and a more efficient ship loading operation.

A number of early studies contributed greatly to the understanding of these problems. They described practical issues related to space allocation, analyzed basic properties related to container handling in storage yards, and proposed formulas to evaluate yard productivity measures such as the total handling effort or the changes in container inventory due to various container-handling strategies (Taleb-Ibrahimi, Castilho, and Daganzo, 1993; Castilho and Daganzo, 1993; Kim, 1997; Chen, 1999).

The *space-allocation* problem is concerned with determining storage locations for containers either individually or as a group. The problem may focus on inbound or outbound container movements, or it may consider both. Storage activities and requirements in yards may be initiated by several sources, such as export containers arriving through gates by truck, at quay by feeder vessels, or by rail service, as well as import containers unloaded from vessels and moved out of the terminal by trucks, feeder vessels, or rail. Each storage activity is characterized by the amount of storage space required, the source and destination of the containers involved (e.g., export containers unloaded from rail have rail as source and the berth as destination), and the starting and ending times of the storage activity. Information regarding these factors can be obtained from delivery schedules or forecasts, as well as unloading/loading schedules for vessels.

We present a formulation that follows that of Kim and Park (2003b) and may be applied to the planning of storage locations for both inbound and outbound containers. The formulation is defined for a planning horizon discretized in T periods, $t = 1, 2, \ldots, T$, according to the starting and ending times of storage activities. Let a_i and b_i represent the starting and the ending times of storage activity $i = 1, 2, \ldots, l$, respectively, and d_i the associated storage space required (i.e., the number of containers in twenty-foot equivalent units). The space allocation problem may then be expressed as a minimum cost multicommodity network flow problem on a time-space network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$. Each storage activity i generates a *source* node S^i and a *destination* node T^i. *Intermediate* nodes j_t identify storage locations $j = 1, 2, \ldots, n$ at time periods t. Arcs from each source node S^i to intermediate nodes in period a_i represent the possible movements of containers from their origin towards the storage locations at the starting time of storage activity i. Symmetrically, arcs from intermediate nodes to a destination node T^i in period b_i stand for the possible movements of containers from the storage locations to their destination at the ending time of storage activity i. The corresponding transportation costs are associated to each of these two types of arcs, but no limits are imposed on their capacity. Holding arcs are defined between intermediate nodes representing the same storage location at consecutive time periods. The capacity u_j of holding arc (j_t, j_{t+1}) corresponds to the stowage capacity of the associated stacking area $j = 1, 2, \ldots, n$, for $t = 1, 2, \ldots, T$. Holding costs are not included in the formulation.

A storage activity i can be described as a route R_{ij} in $\mathcal{G}$ from node S^i to one of the intermediate nodes (storage location), through several holding arcs, to node T^i. The number

of such routes equals the number of available storage locations:

$$R_{ij} : S^i \rightarrow j_{a_i} \rightarrow j_{a_i+1} \rightarrow j_{a_i+2} \rightarrow \cdots \rightarrow j_{b_i-1} \rightarrow j_{b_i} \rightarrow T^i$$

A unit transportation cost c_{ij} is associated to each route R_{ij} of storage activity i. Let $\delta_{R_{ij}}^{j_t}$ equal 1 if arc (j_t, j_{t+1}) is included in route R_{ij} of storage activity i and 0, otherwise.

Define the decision variables x_{ij} as the number of containers of storage activity i that moves following route R_{ij}. The space allocation problem can then be formulated as follows:

$$\text{Minimize} \quad \sum_{i=1}^{l} \sum_{k=1}^{n} c_{ij} f_{ik} \tag{94}$$

$$\text{Subject to} \quad \sum_{i=1}^{l} \delta_{R_{ij}}^{j_t} f_{ik} \leq u_j \qquad j = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, T \tag{95}$$

$$\sum_{j=1}^{n} f_{ik} = d_i \qquad i = 1, 2, \ldots, l \tag{96}$$

$$f_{ik} \geq 0 \qquad \text{and integer} \quad i = 1, 2, \ldots, l, \; j = 1, 2, \ldots, n \tag{97}$$

The objective function (94) minimizes the total transportation cost of all the storage activities. Constraints (95) enforce the storage space limitations at each storage location and time period. Constraints (96) indicate that the space requirement of each storage activity must be satisfied. Notice that storage activities have been defined with constant space requirements for all time periods. A storage activity with a time-varying space requirement may be represented in this formulation as a combination of multiple storage activities with different starting and ending times and space requirement, respectively.

Formulation (94) - (97) corresponds to an integer, capacitated, minimum cost multicommodity flow problem, which is NP-hard (assuming all storage activities begin and end at the same time, the problem reduces to a knapsack problem, which is a well-known NP-hard problem). Moreover, real-world problems are frequently of very large dimensions. Thus, for example, the formulation for a container terminal with 20 blocks and 4 berths yields a mathematical model with more than 100,000 storage activities. Most contributions that focused on the container terminal storage activities thus propose heuristic solution methods. Methodology based on mathematical programming methods for large-scale integer multicommodity network flow problems has been proposed by Barnhart *et al.* (1998) and Barnhart, Hane, and Vance (2000).

Cao and Uebe (1995) addressed the static, single-period version of the previous problem and proposed a transportation problem formulation. Zhang *et al.* (2003) generalized the multi-period problem of Kim and Park (2003b) by including into the objective function terms representing the balancing of handling activities among different storage areas, as well as the

minimization of the total distance traveled by yard trucks between storage areas and quay cranes. The authors considered the flows of both inbound and outbound containers. They decomposed the problem into two stages. The first determined the total number of unloaded (from ships) and received (from land modes) containers to balance their distribution among yard blocks. The second stage determined the number of containers associated to each vessel allocated to each yard block in order to minimize the total travel cost of yard trucks. The first stage problem was formulated as a linear programming model, while the second took the form of a transportation problem.

Several other studies addressed aspects of the storage-space allocation problem different from Kim and Park (2003b). Kim and Bae (1998) examined transportation and dynamic programming models to schedule the remarshaling operations for outbound containers. Kim, Park, and Ryu (2000) addressed the problem of locating individual outbound containers considering the container weights. Kim and Park (2003a) focused on the space-allocation problem for export containers only. They considering various practical constraints and proposed integer programming formulations. Kim and Kim (1999) addressed the space-allocation problem for import containers in which the stacking height and the amount of space (not the storage locations) are simultaneously determined to accommodate dynamically changing space requirements. The authors proposed a non-linear programming model and a solution method based on Lagrangean relaxation. Kozan (2000), Preston and Kozan (2001), and Mettfeld and Kopfer (2003) proposed integer linear programming models to simultaneously determine storage locations and schedules of transfer operations. Castilho and Daganzo (1991) and Holguin-Veras and Jara-Diaz (1999) proposed methods for determining prices of storage spaces for containers or general cargo.

5.5 Allocation and Dispatching of Yard Cranes and Transporters

Yard equipment is an important container terminal resource that includes yard cranes and transporters, i.e., yard trucks and straddle carriers. In automated container terminals, automated guided vehicles belong to this category. The main decision operational issue at this level concerns the assignment of container-handling tasks to different pieces of equipment. Yard equipment is considered less critical than berths, storage spaces, and quay cranes for which schedules are constructed several hours, or even days, in advance. Consequently, no such planning is performed for yard equipment. Instead, decisions on allocation and dispatching yard equipment are made only for the limited number of activities to be carried out in the near future (e.g., one hour and less).

The objectives of allocating and dispatching yard cranes and transporters are usually to minimize the total travel distance, the total waiting time, or the total delays of equipment beyond prespecified due times. The dispatching problem for transporters was studied by Bish (2003), Kim and Bae (1999, 2004), Bose *et al.* (2000), and Grunow *et al.* (2000). Zhang *et al.* (2002), Cheung, Li, and Lin (2002), Kim, Lee, and Hwang (2003), Lai and Lam (1994), and Lai and Leung (1996) addressed the allocation and dispatching problem for yard

cranes.

Consider two ships, sh^- and sh^+, that need unloading and loading, respectively, and that are berthed around the same time so that they can be served by the same set of k vehicles. It is assumed, however, that no ship-to-ship movement is possible. Let $\mathcal{N}^-$ and $\mathcal{N}^+$ denote the sets of containers that will be unloaded (from ship sh^-) and loaded (onto ship sh^+), respectively. Let c_i be container i in $\mathcal{N}^-$. Associated to each container to be loaded onto a ship is its *current storage location* in the yard area, which is known. Each such container will require a *loaded vehicle trip* from its current storage location to the location of ship sh^+. Let L^+ denote the set of current storage locations of the containers in $\mathcal{N}^+$. Symmetrically, a set of *potential storage locations* in the yard area is reserved for the containers of each unloading ship. Set L^- contains the potential storage locations reserved for all containers to be unloaded from sh^-. Each unloaded container will require a *loaded vehicle trip* from the location of ship sh^- to its *selected* storage location. We make the simplifying assumption that sets L^+ and L^- are disjoint, that is, no container in $\mathcal{N}^-$ can be stored in a location currently occupied by a container in $\mathcal{N}^+$. Let $L = L^- \cup L^+$ and W_i be the subset of L^- where c_i can be stacked.

When the destination of a loaded trip differs from the origin of the next loaded trip on a vehicle schedule, the vehicle needs to make an empty, *repositioning* movement. The total amount of empty vehicle trips thus depends on the sequence of loaded trips of each vehicle and should be minimized by, for example, matching each loaded trip for an unloaded container to a loaded trip for a loading container The goal therefore is to (1) determine a storage location for each unloaded container, (2) construct round vehicle trips, and (3) assign round trips to each vehicle. The first problem consists in assigning each unloaded container $c_i \in \mathcal{N}^-$ to a candidate storage location $p \in W_i$. Constructing round vehicle trips, the second problem, corresponds to matching a trip loaded with an unloaded container with a trip loaded with a loading container. Finally, the third problem consists in constructing sequences of round trips and assigning these to vehicles to minimize the make-span.

The following formulation follows the model proposed by Bish (2003) to simultaneously address the first two problems: determines storage locations for containers being unloaded from ships, schedule unloading and loading operations, and dispatches AGVs. The model is based on a network where supply nodes, with unit supply, correspond to unloaded containers $c_i \in \mathcal{N}^-$, demand nodes, with unit demands, correspond to current storage locations of loading containers $l_q, q \in L^+$, and transshipment nodes l_p stand for potential storage locations $p \in W_i$ reserved for unloading containers. The arc set is given by $\mathcal{A} = \{(c_i, l_p) | c_i \in \mathcal{N}^-, p \in W_i\} \cup \{(l_p, l_q) | p \in L^-, q \in L^+\}$, each arc with unit capacity. Arc $(c_i, l_p) \in \mathcal{A}$ corresponds to the trip of an unloading container (c_i that is to be stored in location l_p; Its cost t_{lp} is thus the corresponding travel time. Arc $(l_p, l_q) \in \mathcal{A}$ corresponds to the empty trip of a vehicle that just completed the delivery on an unloaded container to location l_q. The vehicle moves to location l_q to pick up a loading container. The cost of the arc λ_{pq} is thus the travel time of the empty movement.

Define the decision variables $X_{uv}, (u, v) \in \mathcal{A}$, that equal 1 if the activity corresponding

to arc (u, v) is to be performed (the arc is selected) and 0 otherwise. We can now model the problem as a transshipment formulation:

$$\text{Minimize} \quad \sum_{c_i \in \mathcal{N}^-} \sum_{p \in W_i} t_{lp} X_{c_i l_p} + \sum_{p \in L^-} \sum_{q \in L^+} \lambda_{pq} X_{l_p l_q} \tag{98}$$

$$\text{Subject to} \quad \sum_{p \in W_i} X_{c_i l_p} = 1 \qquad c_i \in \mathcal{N}^- \tag{99}$$

$$\sum_{c_i \in N^-} X_{c_i l_p} = \sum_{q \in L^+} X_{l_p l_q} \qquad p \in L^- \tag{100}$$

$$\sum_{p \in L^-} X_{l_p l_q} = 1 \qquad q \in L^+ \tag{101}$$

$$X_{uv} \in \{0, 1\} \qquad \text{for all } (u, v) \in \mathcal{A} \tag{102}$$

The objective function (98) minimizes the total assignment and matching-related travel time of the yard equipment. Constraints (99) ensure that each unloaded container is assigned to exactly one loaded trip and thus, to exactly one unloading location. Constrains (100) enforce the flow-balance requirements at storage sites for unloading containers. Relations (101) ensure that each loaded trip with an unloaded container is matched with a loaded trip with a loading container. The integrality conditions (102) imposed on the decision variables are not really required due to the total unimodularity property the formulation. The author proposed a heuristic for realistically-sized problem instances. The solution of the model yields a set of round trips for vehicles, which consist of a loaded movement with an unloaded container to a storage location, an empty repositioning trip from this location to a pickup location of a loading container, a loaded movement from this pickup location to a loading position under a quay crane, and an empty travel from there to the unloading position of another quay crane. These round trips are then assigned to vehicles by a list-scheduling heuristic.

Several other studies have been dedicated to issues related to yard equipment allocation and dispatching. Kim and Bae (1999) addressed the dispatching problem of AGVs in automated container terminals under the assumption that storage locations of containers as well as the schedules for unloading and loading operations by quay cranes are given. The scope is thus narrower than that of Bish (2003). Kim and Bae (1999) considered quay cranes are the most expensive equipment in container terminals. The model they propose aimed therefore to minimize the total idle time of a quay crane resulting from late arrivals of AGVs as well as the associated total travel time. The authors suggested a network-based mixed-integer linear model and provided a heuristic algorithm. Nodes in the network represent "events" in time and space that correspond to moments (the event time) when vehicles pick up containers. An arc from one node to another indicates that the time lapse between the event times of the two operations allows a vehicle to drop off a container at the destination node after completing its pickup oat the origin node. The problem is then to find the optimal routes on the network, each representing a sequence of delivery tasks assigned to a specific vehicle.

Kim and Bae (2004) extended this approach to the case with multiple quay cranes in which dual cycle operation of AGVs is allowed.

Böse *et al.* (2000) addressed the dispatching problem of straddle carriers with the objective of minimizing the delays of quay cranes. They also assumed that storage locations of containers are already determined. The problem then becomes to assign delivery tasks of unloading and loading containers to straddle carriers. The problem is basically the same as the one addressed by Kim and Bae (2004) except that Böse *et al.* (2000) do not consider the vehicle travel times. The authors did not provide a formal model but proposed a genetic meta-heuristic to search for solutions. Bish *et al.* (2005) represented the dispatching of vehicles as a machine-scheduling problem and proposed a greedy heuristic that assigned deliveries (jobs) to the first available vehicle (machine) which may arrive on time at the designated destination location, i.e., the quay crane or the location of the loading container. The authors showed that the greedy heuristic yields a job sequence that minimizes the makespan. Grunow, Günther, and Lehmann (2004) generalized the previous studies by considering the dispatching problem for multi-load vehicles in automated container terminals. Indeed, while previous studies assumed that a vehicle can move only one container at a time, the authors considered the fact that a vehicle which moves one forty-foot container can move two twenty-foot containers. They proposed a mixed-integer linear programming model and suggested priority rules for dispatching delivery tasks to two-load AGVs.

Similar studies have been done for yard cranes. The objectives considered in most cases were the total waiting time and the total delays of trucks. Zhang *et al.* (2002) and Cheung, Li, and Lin (2002) solved static versions of the crane deployment problem when the total workload at each storage area is known in advance. Zhang *et al.* (2002) proposed a mixed-integer programming model and addressed it by a method based on Lagrangean relaxation. Cheung, Li, and Lin (2002) addressed a similar problem but removed the restriction that crane movements must be completed within a single period. This allows to use a shorter period length resulting in a more accurate model. A successive piecewise-linear approximation method was proposed. Kim, Lee, and Hwang (2003) addressed the problem of sequencing transfer tasks of a yard crane for outside trucks in dynamic situations where new trucks arrive continuously. A dynamic programming model was suggested and decision rules derived by a reinforcement learning technique were proposed. Lai and Lam (1994) and Lai and Leung (1996) proposed various dispatching rules for yard cranes and tested the them by simulation.

6 Strategic Planning of Multimodal Systems

The focus of the models and methods presented in this section is broad: strategic planning issues at the international, national, and regional level, where the movements of several commodities through the transportation networks and services of several carriers are considered simultaneously. The main questions address the evolution of a given transportation system and its response to various modifications in its environment: evolution of the "local" or international socio-economic environment resulting in modifications to the patterns and volumes of production, consumption, and trade; modifications to existing policies and legislation and introduction of new ones (e.g., environment-related taxes); changes to existing infrastructure; variations in energy prices; modifications to labor conditions; carrier mergers; introduction of new technologies, and so on and so forth. These questions are often part of cost-benefit analyzes and comparative studies of policy and investment alternatives.

A strategic planning methodology identifies and represents the fundamental components of a transportation system – demand, supply, performance measures and decision criteria – and their interactions. It models flow volumes by commodity and transportation mode, as well as associated performance measures, defined on a network representation of the transportation system. It aims to achieve a sufficiently good simulation of the global behavior of the system to offer a correct representation of the current situation and serve as an adequate analysis tool for planned or forecast scenarios and policies. It must be tractable and produce results that are easily accessible. No single formulation may address such a broad scope. Consequently, a strategic planning methodology is typically a set of models and procedures. Other than data manipulation tools (e.g., collection, fusion, updating, validation, etc.) and result analysis capabilities (e.g., cost-benefit, environmental impacts, energy consumption policies, etc.), the main components are: 1) *Supply modeling* to represent the transportation modes, infrastructure, carriers, services, and lines; vehicles and convoys; terminals and intermodal facilities; capacities and congestion; economic, service, and performance measures and criteria. 2) *Demand modeling* to capture the product definitions, identify producers, shippers, and intermediaries, and represent production, consumption, and zone-to-zone (region-to-region) distribution volumes, as well as *mode choices* for transportation. Relations of demand and mode choice to the performance of economic policies and transportation system performance are also addressed here. 3) *Assignment* of multicommodity flows (from the demand model) to the multimode network (the supply representation). This procedure simulates the behavior of the transportation system and its output forms the basis for the strategic analyses and planning activities. The assignment methodology must therefore be both precise in reproducing current situations and sufficiently general to produce robust analyzes of future scenarios based on forecast data.

The prediction of multicommodity freight flows over a multimodal network is an important component of transportation science and has generated significant interest in recent years. However, perhaps due to the inherent difficulty and complexity of such problems, the study of freight flows at the national or regional level has not yet achieved full maturity, in contrast to passenger transportation where the prediction of car and transit flows over

multimodal networks has been studied extensively and several of the research results have been transferred to practice (Florian and Hearn, 1995; Cascetta, 2001). In the following, we review the most frequently used methodologies for freight planning and briefly describe associated references.

Demand

The modeling of demand attempts to describe the economic activities of a region, its production, consumption, import and export of goods. For planning purposes, the output of demand models is a series of product (or commodity group) specific demand matrices indicating the volumes to be moved from one zone to another. The process is often completed by the modeling of mode choice, which specifies for each product and origin-destination combination on what set of transportation infrastructure or services the demand may be moved.

A number of countries have developed *input/output* models of their economy that serve to determine the basic production and attraction of goods (Isard 1951; Casceta 2001 and references within). In order to use an input/output model as a demand model, it is necessary to disaggregate the input/output model inputs and outputs by region and zone. This process is complex, and is usually not integrated with a supply representation and assignment procedure. When an input/output model is not available, the initial determination of origin-destination matrices is carried out by using national statistics on production, consumption, imports and exports combined with surveys of particular industrial sectors to complete missing or unreliable information. This process may be tedious since one has to reconcile data from several sources that may be collected by using different geographical subdivisions or inconsistent product definitions. The results of the disaggregated input/output model or the ad-hoc estimation procedures serve for the initial computation of origin-destination matrices for each product but without a subdivision by mode.

A second class of models that is well studied for the prediction of interregional commodity flows is the *spatial price equilibrium model* and its variants (Friesz, Tobin, and Harker, 1983; Harker and Friesz 1986a,b; and Harker 1987; see also Florian and Hearn, 1995 and Nagurney, 1993). This class of models determines simultaneously the flows between producing and consuming regions, as well as the selling and buying prices that satisfy the spatial equilibrium conditions. A spatial equilibrium is reached when for all pairs of supply and demand regions with a positive commodity flow, the unit supply price plus the unit transportation cost is equal to the unit demand price; the sum is larger than this price for all pairs of regions with no exchanges. A simple network (bipartite graph) is generally used to represent the transportation system. These models rely to a large extent on the supply and demand functions of producers and consumers, respectively, which are rarely available and quite difficult to calibrate. There are relatively few applications of this class of models for the determination of demand by product and these deal with specific commodities such as crude oil, coal or milk products.

Mode Choice

Mode-choice models aim to describe the set of transportation modes or services that may be used to carry specific products of groups thereof. The mode-choice definition may be rather general, e.g., petroleum moves by ship and pipeline, extremely specific indicating a particular set of single or multimodal paths for a given product, shipper, and origin-destination pair, or anywhere in between. The level of detail of modal specification need not be the same for all products or inter-zonal trade flows. The specification of mode choice may be inferred from historical data and shipper surveys or it may result from a formal description and modeling effort (Winston 1983). The output of this process are either coefficients that indicate how to split the demand of a given origin-destination pair between the paths of a given set, or origin-destination demand matrices with particular sets of allowed modes.

Random utility models, developed and largely used for the analysis and planning of person transportation systems, have been proposed for freight transportation as well but their use in actual applications is scarce (Cascetta 2001). The huge number of paths that have to be explicitly generated and stored, coupled to the challenge of performing this task for forecast data, may explain this phenomenon. At aggregated levels, mode choices have been specified for particularly important product flows by explicitly surveying the major logistic chains used between pairs of macro regions.

Supply Representation and Assignment

Once modal origin-destination matrices have been developed by some means, the next step is to assign them to the supply network model by using some route choice mechanism. The results of such an assignment model - product flows and performance measures - form part of the input to demand and cost-benefit modeling and analysis.

One class of assignment mechanisms is based again on the application of *random utility models* to the choice of paths defined previously by the mode-choice phase. It is noteworthy that the attributes of pre-defined paths are determined by the state of the network at generation time and are not responsive to assignment results. Thus, for example, congestion conditions are very difficult to represent. Moreover, the utility and choice models have to be calibrated, and all paths have to be generated, for each scenario, which is quite difficult to perform when forecast data is used.

Another class, *network optimization models* enable the prediction of multicommodity flows over a multimodal network that represents the transportation facilities at a level of detail appropriate for a nation or region. The demand and mode choice are exogenous and intermodal shipments are permitted. Within the specified mode choice, the optimization (assignment) engine determines the best (with respect to the specified network performance measures) multimodal paths for each product and origin-destination pair. The emphasis is on the proper representation of the network and its different transportation modes, the corresponding intermodal transfer operations, the various criteria used to determine the movement of freight, the interactions and competition for limited resources captured through the representation of congestion effects, and the associated estimation of the traffic distribution over the transportation system considered to be used for comparative studies or for discrete

time multi-period analyses.

Studies in the 1970's used rather simple network representations (e.g., Jones and Sharp, 1977; Sharp, 1979). Several studies also attempted to extend spatial equilibrium models to include more refined network representations and to consider congestion effects and shipper-carrier interactions. Friesz, Gottfried, and Morlok (1986) present a sequential model which uses two network representations: detailed separate networks for each carrier, and an aggregate, shipper-perceived network. On each carrier network commodities are transported at the least total cost. On the shipper-perceived network, traffic equilibrium principles are used to determine the carriers that shippers choose to move their traffic. This approach was quite successful in the study of logistics of products where a very limited number of shippers and carriers interact and strongly determine the behavior of the system. A typical example is the coal market between electric utilities in the United States and their suppliers in exporting countries. Friesz and Harker (1985), Harker and Friesz (1986a,b), Harker (1987, 1988), and Hurley and Petersen (1994) present more elaborate formulations. This line of research has not, however, yielded practical planning models and tools yet, mainly because the formulations become too large and complex when applied to realistic situations. For a more detailed review of these efforts see Guélat, Florian, and Crainic (1990) and Crainic, Florian, and Léal (1990).

Models based on more sophisticated representations of the supply network were introduced by Guélat, Florian, and Crainic (1990) and Jourquin and Beuthe (1996). We follow the former in the following presentation. The proposed modeling framework is that of a multimodal network, made up of modes, nodes, links, and intermodal transfers, on which multiple products are to be moved by specific vehicles and convoys between given origin and destination points. Here, a mode is a means of transportation having its own characteristics, such as vehicle type and capacity, as well as specific cost measures. Depending on the scope and level of detail of the strategic study, a mode may represent a carrier or part of its network representing a particular transportation service, an aggregation of several carrier networks, or specific transportation infrastructures such as ports.

The network consists of nodes $\mathcal{N}$, links $\mathcal{A}$, modes $\mathcal{M}$, and transfers $\mathcal{T}$ that represent all possible physical movements on the available infrastructure. To capture the modal characteristics of transportation, a link $a \in \mathcal{A}$ is defined as a triplet (i, m, j), where $i \in \mathcal{N}$ is the origin node, $j \in \mathcal{N}$ is the destination node, and $m \in \mathcal{M}$ is the mode. Parallel links are used to represent situations where more than one mode is available for transporting goods between two adjacent nodes. This network representation enables easy identification of the flow of goods by mode, as well as various cost functions (e.g. operating cost, time delay, energy consumption, emissions, noise, risk, etc.) by product and mode. To model intermodal shipments, one must allow for mode transfers at certain nodes of the network and compute the associated costs and delays. Intermodal transfers $t \in \mathcal{T}$ at a node of the network are modeled as link to link, hence mode to mode, allowed movements. A path in this network then consists of a sequence of directed links of a mode, a possible transfer to another mode, a sequence of directed links of the second mode, and so on.

A product is any commodity (or collection of similar products) – goods or passengers – that generates a link flow. Each product $p \in \mathcal{P}$ transported over the multimodal network is shipped from certain origins $o \in \mathcal{N}$ to certain destinations $d \in \mathcal{N}$ within the network. The demand for each product is exogenous and is specified by a set of O-D matrices. The mode choice for each product is also exogenous and is indicated by defining for each O-D matrix a subset of modes allowed for transporting the corresponding demand. Shipper behavior is assumed to be reflected in these O-D matrices and associated mode choice. Let $g^{m(p)}$ be a demand matrix associated with product $p \in \mathcal{P}$, where $m(p) \subseteq \mathcal{M}$ is the subset of modes that may be used to move this particular part of product p.

The flows of product $p \in \mathcal{P}$ on the multimodal network are the decision variables of the model. Flows on links $a \in \mathcal{A}$ are denoted by v_a^p and flows on transfers $t \in \mathcal{T}$ are denoted by v_t^p; v stands for the vector of all product flows. Vehicle and convoy (e.g., train) movements are deduced from these flows. Cost functions are associated with the links and transfers of the network. For product p, the respective average cost functions $s_a^p(v)$ and $s_t^p(v)$ depend on the transported volume of goods. Then, the total cost of product p on arc a is $s_a^p(v)v_a^p$, and it is $s_t^p(v)v_t^p$ on transfer t. The total cost over the multimodal network is the function F, which is to be minimized over the set of flow volumes that satisfy the flow conservation and non negativity constraints:

$$F = \sum_{p \in P} \left(\sum_{a \in \mathcal{A}} s_a^p(v) v_a^p + \sum_{t \in \mathcal{T}} s_t^p(v) v_t^p \right). \tag{103}$$

Let $\mathcal{L}_{od}^{m(p)}$ denote the set of paths that for product p lead from origin o to destination d using only modes in $m(p)$. The path formulation of the flow conservation equations are then:

$$\sum_{l \in \mathcal{L}_{od}^{m(p)}} h_l = g_{od}^{m(p)} \qquad o,\ d \in \mathcal{N},\ p \in \mathcal{P},\ m(p) \subseteq \mathcal{M}, \tag{104}$$

where h_l is the flow on path $l \in \mathcal{L}_{od}^{m(p)}$. These constraints specify that the total flow moved over all the paths that may be used to transport product p must be equal to the demand for that product. The non negativity constraints are:

$$h_l \geq 0, \qquad l \in \mathcal{L}_{od}^{m(p)},\ o,\ d \in \mathcal{N},\ p \in \mathcal{P},\ m(p) \subseteq \mathcal{M}. \tag{105}$$

The relation between arc flows and path flows is $v_a^p = \sum_{l \in \mathcal{L}^p} \delta_{al} h_l,\ a \in \mathcal{A},\ p \in \mathcal{P}$, where $\mathcal{L}^p$ is the set of all paths that may be used by product p, and $\delta_{al} = 1$ if $a \in l$ (and 0, otherwise) is the indicator function which identifies the arcs of a particular path. Similarly, the flows on transfers are $v_t^p = \sum_{l \in \mathcal{L}^p} \delta_{tl} h_l,\ t \in \mathcal{T},\ p \in \mathcal{P}$, where $\delta_{al} = 1$ if $t \in l$ (and 0, otherwise). Then, the system optimal multiproduct, multimodal assignment model consists of minimizing (103), subject to constraints (104) and (105). The optimality principle ensures that in the final flow distribution, for each product, demand matrix, and origin-destination pair, all paths with positive flows will have the same marginal cost (lower than on the other

paths). The algorithm developed for this problem exploits the natural decomposition by product and results in a Gauss-Seidel-like procedure which allows the solution of large size problems in reasonable computational times (Guélat, Florian, and Crainic, 1990).

This model and algorithm are embedded in the STAN interactive-graphic system where they are complemented by a large number of tools to input, display, analyze, modify, and output data, as well as implement demand, mode choice, performance, and analysis models. See Larin *et al.* (2000) for a detailed description of the STAN system, components, interfaces, and tools. STAN has been applied successfully in practice for scenario analysis and planning, and several agencies and organizations in a number of countries around the world use it (Crainic, Florian, and Léal, 1990; Guélat, Florian, and Crainic, 1990; Crainic *et al.*, 1990; Crainic, Florian, and Larin, 1994; and Crainic *et al.*, 1999, 2002).

7 Perspectives

Intermodal transportation, particularly container-based, is steadily growing and will continue to do so in the foreseeable future. This is accompanied by the evolution of the regulatory, economic, and technological environment of the industry. Enhanced planning and management procedures and decision technologies are thus required, offering both great opportunities and significant challenges for the Operations Research community. On a general note, while many significant methodological advances have been achieved and several have been successfully transferred to actual practice, many problems have received scant attention. Moreover, advances in vehicle, infrastructure, and communication technologies yield new problems and require that problems already studied be revisited. In this section, we identify some of these trends, opportunities, and challenges.

Container terminals, mainly located in ports, are a case in point. Most research dedicated to this area is very recent and aimed at operational issues. This may be explained by the fact that terminals are often seen as bottlenecks in freight transportation and efforts are therefore dedicated to improving their efficiency and productivity. The development of comprehensive models for strategic and tactical planning of container terminals offers significant research opportunities. Moreover, the trend one observes in container terminal automation makes this research direction extremely timely. How to represent automatic operations in planning models raises interesting questions, however. The automation of container terminals also opens up research opportunities in real-time decision and control of operations. Automated equipment collects and transmits data in real-time. This data, together with historical information and the plan of operations, could be used to automate and, in some case nearly optimize, real-time decisions. New models are required, as well as appropriate solution methods. Automation of terminals also requires revisiting operational planning models.

Compared to terminals, more work has been dedicated to carrier strategic and tactical planning issues. Yet, new problems emerge and many challenging research opportunities exist. Enhancing the models to better represent operation characteristics and to better integrate line and terminal activities is such an opportunity; integrating a representation of resource (vehicles, power, manpower) circulation and scheduling is another. The planning and operations of the "new" rail intermodal-service networks, operating regular and fixed services on a full-asset-utilization basis and enforcing advance bookings, define new challenging problems for Operations Research and Transportation Science.

The more comprehensive integration of the time-dependency of decisions and of the stochasticity of data and operations into strategic/tactic models is a major research challenge and opportunity. Indeed, many models aimed at strategic and tactical planning issues are static and almost all are deterministic. The conventional wisdom seems to be that such models plan based on "average" forecast data, while actual operations provide the "recourse" to adjust the plan to the day-to-day reality. Are we missing something? Could we build better plans and operate more efficiently by taking stochasticity into account at planning level? These and similar questions open up a fascinating research field.

Many operational problems are represented as time-dependent formulations. Crew scheduling, for example. Some, such as the management of empty vehicles, are also cast as stochastic formulations. Research opportunities exist in the development and enhancement of such formulations in many sectors of the industry. Thus, for example, very few results have been reported relative to the container fleet management problem. Moreover, each of these efforts was dedicated to one part of the system only. No comprehensive model of container fleet management on land and ocean is known to the authors. Moreover, while the scheduling and operation of various resources – vehicles, traction units, crews, etc. – are clearly related, most current formulations consider them separately. Research is required in this area.

The growth in the deployment of *Intelligent Transportation Systems (ITS)* and the electronic society will continue to impact the planning and operations of freight transportation. ITS and e-business technologies and procedures increase the flow of data, improve the timeliness and quality of information, and offer the possibility to control and coordinate operations in real-time. Research is required to adequately model the various planning and management problems under ITS and real-time information and to develop efficient solution methods. These efforts must target carriers, terminals, as well as the entire intermodal chain. The scheduling, assignment, dispatching, routing, and re-routing of equipment are obvious and challenging subjects. As important is the impact on planning. Consider, to illustrate, the uncertainties related to the exact manifest of arriving container ships, their exact arrival time, and the destination (and, eventually, the carrier contracted for the next leg of the journey) of each container. ITS technology will deliver precise information earlier thus significantly reducing the uncertainty for the managers of the terminal and of the carriers that are next in the intermodal chain. Will uncertainty disappear completely? Unlikely. But its representation in planning and operations models will change.

Very few efforts have been dedicated to the intermodal chain. The e-business environment forces the issue and ITS offers the technological support. This largely unexplored field offers numerous research opportunities and challenges. The coordination (synchronization, in some cases) of plans and operations of independently owned or managed carriers and terminals is such a case. The uncertainties related to the operation of each element of the chain, the relations among these uncertainties, as well as their propagation within the intermodal chain are of prime importance in this context and pose considerable modeling and algorithmic challenges.

Recent years have brought to the forefront security issues related to transportation, ports, and border crossing. Planning and operations models and methods must be revisited and new ones must be proposed to address these issues, for each participant in the intermodal chain, as well as for the entire chain.

Most problems mentioned in this Chapter are NP-Hard and the formulations proposed are large-scale, mixed-integer combinatorial models. Stochastic, time-dependent formulations make resolution efforts even more difficult. And the need to build more comprehensive models is not making them any easier to solve. Significant research must thus be dedicated to the methodological aspects, including the study of models to develop stronger formulations

and bounds. From an algorithmic point of view, the profession must continue to aim for powerful exact methods to address continuously larger problem instances. The dimension and complexity of the problems we face continuously overcome the capability of exact solution methods, however. Approximate solution methods, meta-heuristics in particular, thus play an increasingly important role in obtaining good solutions to difficult problems within reasonable computing times. Much work is still needed to develop more efficient and robust procedures and to better understand the conditions under which each method performs best. Solution methods that combine characteristics of two or more meta-heuristics in a sequential or parallel computing setting offer interesting, but challenging perspectives.

Parallel and distributed computation indeed offers interesting perspectives with potentially great rewards: to solve realistically modeled and dimensioned problem instances within reasonable times. Each class of problems and algorithms presents its own challenges. Promising research areas are the parallel exploration of branch-and-bound trees, the collaborative search undertaken by several meta-heuristics or by meta-heuristics and exact methods, and the development of hierarchical methods that combine different parallel models and methods (e.g., a first level of cooperating meta-heuristics and branch-and-bound that each call, at a lower level, parallel procedures to evaluate solutions or bounds). Advanced decomposition techniques are also required, particularly related to the resolution of time-dependent problems. Parallel computing offers the possibility to design solution architectures to efficiently address complex requests in real, or quasi-real time. These ideas have just begun to be considered, but present great potential for the development of intelligent and efficient decision support tools for real-time intermodal transportation systems.

Acknowledgments

Funding for this project has been provided by the Natural Sciences and Engineering Council of Canada, through its Discovery Grant program. This work was supported in part by the Regional Research Centers Program (Research Center for Logistics Information Technology) granted by the Korean Ministry of Education & Human Resources Development.

While working on this project, Dr. T.G. Crainic was Adjunct Professor at the Département d'informatique et de recherche opérationnelle of the Université de Montréal (Canada) and at Molde University College (Norway).

References

Addinnour-Helm, S. and Venkataramanan M.A. (1998). Solution Approaches to Hub Location Problems. *Annals of Operations Research*, 78:31–50.

Ahuja, R.K., Magnanti, T.L., and Orlin, J.B. (1993). *Network Flows – Theory, Algorithms, and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

Aikens, C.H. (1985). Facility Location Models for Distribution Planning. *European Journal of Operational Research*, 22:263–279.

Alessandri, A., Sacone, S., and Siri, S. (2006). Modelling and Receding-Horizon Control of Maritime Container Terminals. *Journal of Mathematical Modelling and Algorithms*. forthcoming.

Armacost, A.P., Barnhart, C., and Ware, K.A. (2002). Composite Variable Formulations for Express Shipment Service Network Design. *Transportation Science*, 36(1):1–20.

Armacost, A.P., Barnhart, C., Ware, K.A., and Wilson, A.M. (2004). UPS Optimizes Its Air Network. *Interfaces*, 34(1):15–25.

Assad, A.A. (1980). Models for Rail Transportation. *Transportation Research Part A: Policy and Practice*, 14A:205–220.

Aykin, T. (1990). On a Quadratic Integer Program for the Location of Interacting Hub Facilities. *European Journal of Operational Research*, 46(3):409–411.

Aykin, T. (1994). Lagrangian Relaxation Based Approaches to Capacitated Hub-and-Spoke Network Design Problem. *European Journal of Operational Research*, 79(3):501–523.

Aykin, T. (1995a). The Hub Location and Routing Problem. *European Journal of Operational Research*, 83(1):200–219.

Aykin, T. (1995b). Networking Policies for Hub-and-Spoke Systems with Application to the Air Transportation System. *Transportation Science*, 29(3):201–221.

Balakrishnan, A., Magnanti, T.L., and Mirchandani, P. (1997). Network Design. In Dell'Amico, M., Maffioli, F., and Martello, S., editors, *Annotated Bibliographies in Combinatorial Optimization*, pages 311–334. John Wiley & Sons, New York, NY.

Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors (1995). *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*. North-Holland, Amsterdam.

Barnhart, C., Hane, C.A., and Vance, P.H. (2000). Using Branch-and-Price-and-Cut to Solve Origin-Destination Integer Multicommodity Flow Problems. *Operations Research*, 48(2):318–326.

Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W.F., and Vance, P.H. (1998). Branch-and-Price: Column Generation for Solving Huge Integer Programs. *Operations Research*, 46(3):316–329.

Barnhart, C. and Schneur, R.R. (1996). Network Design for Express Freight Service. *Operations Research*, 44(6):852–863.

Bish, E.K. (2003). A Multiple-Crane-Constrained Scheduling Problem in a Container Terminal. *European Journal of Operational Research*, 144(1):83–107.

Bish, E.K., Chen, F.Y., Leong, Y.T., Nelson, B.L., Ng, J.W.C., and Simchi-Levi, D. (2005). Dispatching Vehicles in a Mega Container Terminal. *OR Spectrum*, 27(4):491–506.

Böse, J., Reiners, T., Steenken, D., and Voß, S. (2000). Vehicle Dispatching at Seaport Container Terminals Using Evolutionary Algorithms. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, volume 2, page 2025. IEEE Computer Society. 10 pages.

Bostel, N. and Dejax, P. (1998). Models and Algorithms for Container Allocation Problems on Trains in a Rapid Transshipment Shunting Yard. *Transportation Science*, 32(4):370–379.

Bourbeau, B., Gendron, B., and Crainic, T.G. (2000). Branch-and-Bound Parallelization Strategies Applied to a Depot Location and Container Fleet Management Problem. *Parallel Computing*, 26(1):27–46.

Braklow, J.W., Graham, W.W., Hassler, S.M., Peck, K.E., and Powell, W.B. (1992). Interactive Optimization Improves Service and Performance for Yellow Freight System. *Interfaces*, 22(1):147–172.

Brown, G.G., Cormican, K.J., Lawphongpanich, S., and Widdis, D.B. (1997). Optimizing Submarine Berthing with a Persistence Incentive. *Naval Research Logistics*, 44:301–318.

Brown, G.G., Lawphongpanich, S., and Thurman, K.P. (1994). Optimizing Ship Berthing. *Naval Research Logistics*, 41:1–15.

Buedenbender, K., Grünert, T., and Sebastian, H.-J. (2000). A Hybrid Tabu Search/Branch and Bound Algorithm for the Direct Flight Network Design Problem. *Transportation Science*, 34(4):364–380.

Campbell, J.F. (1994a). Integer Programming Formulations of Discrete Hub Location Problems. *European Journal of Operational Research*, 72:387–405.

Campbell, J.F. (1994b). A Survey of Network Hub Location Problems. *Studies in Locational Analysis*, 6:31–49.

Campbell, J.F. (1996). Hub Location and the p-Hub Median Problem. *Operations Research*, 44(6):923–935.

Campbell, J.F., Ernst, A.T., and Krishnamoorthy, M. (2002). Hub Location Problems. In Drezner, Z. and Hamacher, H., editors, *Facility Location: Application and Theory*, pages 373–407. Springer Verlag, Berlin.

Cao, B. and Uebe, G. (1995). Solving Transportation Problems with Nonlinear Side Constraints with Tabu Search. *Computers & Operations Research*, 22(6):593–603.

Cascetta, E. (2001). *Transportation Systems Engineering: Theory and Methods*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Castilho, B.D. and Daganzo, C.F. (1991). Optimal Pricing Policies for Temporary Storage at Ports. *Transportation Research Record*, 1313:66–74.

Castilho, B.D. and Daganzo, C.F. (1993). Handling Strategies for Import Containers at Marine Terminals. *Transportation Research Part B: Methodology*, 27B(2):151–166.

Chen, T. (1999). Yard operations in the Container Terminal - A Study in the Unproductive Moves. *Maritime Policy Management*, 26(1):27–38.

Cheung, R.K. and Chen, C.-Y. (1998). A Two-Stage Stochastic Network Model and Solution Methods for the Dynamic Empty Container Allocation Problem. *Transportation Science*, 32(2):142–162.

Cheung, R.K., Li, C.L., and Lin, W. (2002). Interblock Crane Deployment in Container Terminals. *Transportation Science*, 36(1):79–93.

Christiansen, M., Fagerholt, K., Nygreen, B., and Ronen, D. (2006). Maritime Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, Handbooks in Operations Research and Management Science. North-Holland, Amsterdam.

Christiansen, M., Fagerholt, K., and Ronen, D. (2004). Ship Routing and Scheduling: Status and Perspectives. *Transportation Science*, 38(1):1–18.

Cojeen, H.P. and Dyke, P.V. (1976). The Automatic Planning and Sequencing of Containers for Containership Loading and Unloading. In Pitkin, M., Roche, J.J., and Williams, T.J., editors, *Ship Operation Automation, II: Proceedings of the 2nd IFAC/IFIP Symposium*, pages 415–423. North-Holland Publishing Co., Amsterdam.

Cordeau, J.-F., Laporte, G., Legato, P., and Moccia, L. (2005). Models and Tabu Search Heuristics for the Berth Allocation Problem. *Transportation Science*. forthcoming.

Cordeau, J.-F., Laporte, G., Savelsbergh, M.W.P., and Vigo, D. (2006). Vehicle Routing. In Barnhart, C. and Laporte, G., editors, *Transportation*, Handbooks in Operations Research and Management Science. North-Holland, Amsterdam. forthcoming.

Cordeau, J.-F., Toth, P., and Vigo, D. (1998). A Survey of Optimization Models for Train Routing and Scheduling. *Transportation Science*, 32(4):380–404.

Crainic, T.G. (1988). Rail Tactical Planning: Issues, Models and Tools. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics*, pages 463–509. Springer-Verlag, Berlin.

Crainic, T.G. (2000). Network Design in Freight Transportation. *European Journal of Operational Research*, 122(2):272–288.

Crainic, T.G. (2003). Long-Haul Freight Transportation. In Hall, R.W., editor, *Handbook of Transportation Science*, pages 451–516. Kluwer Academic Publishers, Norwell, MA, second edition.

Crainic, T.G., Dejax, P.J., and Delorme, L. (1989). Models for Multimode Multicommodity Location Problems with Interdepot Balancing Requirements. *Annals of Operations Research*, 18:279–302.

Crainic, T.G. and Delorme, L. (1995). Dual-Ascent Procedures for Multicommodity Location-Allocation Problems with Balancing Requirements. *Transportation Science*, 27(2):90–101.

Crainic, T.G., Delorme, L., and Dejax, P.J. (1993a). A Branch-and-Bound Method for Multicommodity Location with Balancing Requirements. *European Journal of Operational Research*, 65(3):368–382.

Crainic, T.G., Dufour, G., Florian, M., and Larin, D. (1999). Path Analysis in STAN. In C. Zopounidis and D. Despotis, editors, *Proceedings 5th International Conference of the Decision Sciences Institute*, pages 2060–2064. New Technologies Publications, Athens, Greece.

Crainic, T.G., Dufour, G., Florian, M., and Larin, D. (2002). Path Recovery/Reconstruction and Applications in Nonlinear Multimodal Multicommodity Networks. In Gendreau, M. and Marcotte, P., editors, *Transportation and Network Analysis: Current Trends Miscellanea in Honor of Michael Florian*, pages 2060–2064. Kluwer Academic Publishers, Norwell, MA.

Crainic, T.G., Ferland, J.-A., and Rousseau, J.-M. (1984). A Tactical Planning Model for Rail Freight Transportation. *Transportation Science*, 18(2):165–184.

Crainic, T.G., Florian, M., Guélat, J., and Spiess, H. (1990a). Strategic Planning of Freight Transportation: STAN, An Interactive-Graphic System. *Transportation Research Record*, 1283:97–124.

Crainic, T.G., Florian, M., and Larin, D. (1994). STAN: New Developments. In Al S. Khade and R. Brown, editors, *Proceedings of the 23rd Annual Meeting of the Western Decision Sciences Institute*, pages 493–498. School of Business Administration, California State University, Stanislaus, CA.

Crainic, T.G., Florian, M., and Léal, J.-E. (1990b). A Model for the Strategic Planning of National Freight Transportation by Rail. *Transportation Science*, 24(1):1–24.

Crainic, T.G., Gendreau, M., and Dejax, P.J. (1993b). Dynamic Stochastic Models for the Allocation of Empty Containers. *Operations Research*, 41(1):102–302.

Crainic, T.G., Gendreau, M., Soriano, P., and Toulouse, M. (1993c). A Tabu Search Procedure for Multicommodity Location/Allocation with Balancing Requirements. *Annals of Operations Research*, 41:359–383.

Crainic, T.G. and Laporte, G. (1997). Planning Models for Freight Transportation. *European Journal of Operational Research*, 97(3):409–438.

Crainic, T.G. and Rousseau, J.-M. (1986). Multicommodity, Multimode Freight Transportation: A General Modeling and Algorithmic Framework for the Service Network Design Problem. *Transportation Research Part B: Methodological*, 20:225–242.

Crainic, T.G. and Roy, J. (1988). O.R. Tools for Tactical Freight Transportation Planning. *European Journal of Operational Research*, 33(3):290–297.

Crainic, T.G., Toulouse, M., and Gendreau, M. (1995a). Parallel Asynchronous Tabu Search for Multicommodity Location-Allocation with Balancing Requirements. *Annals of Operations Research*, 63:277–299.

Crainic, T.G., Toulouse, M., and Gendreau, M. (1995b). Synchronous Tabu Search Parallelization Strategies for Multicommodity Location-Allocation with Balancing Requirements. *OR Spektrum*, 17(2/3):113–123.

Crainic, T.G., Toulouse, M., and Gendreau, M. (1997). Towards a Taxonomy of Parallel Tabu Search Algorithms. *INFORMS Journal on Computing*, 9(1):61–72.

Croxton, K.L., Gendron, B., and Magnanti, T.L. (2003). Models and Methods for Merge-in-Transit Operations. *Transportation Science*, 37(1):1–22.

Croxton, K.L., Gendron, B., and Magnanti, T.L. (2004). Variable Disaggregation in Network Flow Problems with Picewise Linear Costs. Publication CRT-2004-05, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Daganzo, C.F. (1989a). Crane Productivity and Ship Delay in Ports. *Transportation Research Record 1251.*

Daganzo, C.F. (1989b). The Crane Scheduling Problem. *Transportation Research Part B: Methodology*, 23B(3):159–175.

Daganzo, C.F. (1990). The Productivity of Multipurpose Seaport Terminals. *Transportation Science*, 24(3):205–216.

Daganzo, C.F. (2005). *Logistics Systems Analysis.* Springer, Berlin. 4th Edition.

Daskin, M.S. (1995). *Network and Discrete Location. Models, Algorithms, and Applications.* John Wiley & Sons, New York, NY.

Daskin, M.S. and Owen, S.H. (2003). Location Models in Transportation. In Hall, R.W., editor, *Handbook of Transportation Science*, pages 321–371. Kluwer Academic Publishers, Norwell, MA, second edition.

Dejax, P.J. and Crainic, T.G. (1987). A Review of Empty Flows and Fleet Management Models in Freight Transportation. *Transportation Science*, 21(4):227–247.

Delorme, L., Roy, J., and Rousseau, J.-M. (1988). Motor-Carrier Operation Planning Models: A State of the Art. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics*, pages 510–545. Springer-Verlag, Berlin.

Drezner, Z., editor (1995). *Facility Location. A Survey of Applications and Methods.* Springer-Verlag, New York, NY.

Drezner, Z. and Hamacher, H., editors (2002). *Facility Location: Application and Theory.* Springer Verlag, Berlin.

Dror, M., editor (2000). *Arc Routing: Theory, Solutions and Applications.* Kluwer Academic Publishers, Norwell, MA.

Ebery, J., Krishnamoorthy, M., Ernst, A., and Boland, N. (2000). The Capacitated Multiple Allocation Hub Location Problem: Formulations and Algorithms. *European Journal of Operational Research*, 120(3):614–631.

Equi, L., Gallo, G., Marziale, S., and Weintraub, A. (1997). A Combined Transportation and Scheduling Problem. *European Journal of Operational Research*, 97(1):94–104.

Erlenkotter, D. (1978). A Dual–Based Procedure for Uncapacitated Facility Location. *Operations Research*, 26(6):992–1009.

Ermol'ev, Y.M., Krivets, T.A., and Petukhov, V.S. (1976). Planning of Shipping Empty Seaborne Containers. *Cybernetics*, 12:644–646.

Ernst, A.T. and Krishnamoorthy, M. (1996). Efficient Algorithms for the Uncapacitated Single Allocation p-Hub Median Problem. *Location Science*, 4(3):139–154.

European Conference of Ministers of Transport (1986). Terminology for Combined Transport. http:/www.cordis.lu/transport.

Farvolden, J.M. and Powell, W.B. (1991). A Dynamic Network Model for Less-Than-Truckload Motor Carrier Operations. Working Paper 90-05, Department of Industrial Engineering, University of Toronto, Toronto, ON, Canada.

Farvolden, J.M. and Powell, W.B. (1994). Subgradient Methods for the Service Network Design Problem. *Transportation Science*, 28(3):256–272.

Farvolden, J.M., Powell, W.B., and Lustig, I.J. (1992). A Primal Partitioning Solution for the Arc–Chain Formulation of a Multicommodity Network Flow Problem. *Operations Research*, 41(4):669–694.

Florian, M. and Hearn, D. (1995). Network Equilibrium Models and Algorithms. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, pages 485–550. North-Holland, Amsterdam.

Fratar, T.J., Goodman, A.S., and Brant, Jr. A.E. (1960). Prediction of Maximum Practical Berth Occupancy. *Journal of the Waterways and Harbors Division*, pages 69–78.

Friesz, T.L., Gottfried, J.A., and Morlok, E.K. (1986). A Sequential Shipper-Carrier Network Model for Predicting Freight Flows. *Transportation Science*, 20:80–91.

Friesz, T.L. and Harker, P.T. (1985). Freight Network Equilibrium: A Review of the State of the Art. In A.F. Daughety, editor, *Analytical Studies in Transport Economics*, chapter 7. Cambridge University Press, Cambridge.

Friesz, T.L., Tobin, R.L., and Harker, P.T. (1983). Predictive Intercity Freight Network Models. *Transportation Research Part A: Policy and Practice*, 17A:409–417.

Gendron, B. and Crainic, T.G. (1995). A Branch-and-Bound Algorithm for Depot Location and Container Fleet Management. *Location Science*, 3(1):39–53.

Gendron, B. and Crainic, T.G. (1997). A Parallel Branch-and-Bound Algorithm for Multicommodity Location with Balancing Requirements. *Computers & Operations Research*, 24(9):829–847.

Gendron, B., Potvin, J.-Y., and Soriano, P. (1999). Tabu Search with Exact Neighbor Evaluation for Multicommodity Location with Balancing Requirements. *INFOR*, 37(3):255–270.

Gendron, B., Potvin, J.-Y., and Soriano, P. (2003a). A Parallel Hybrid Heuristic for the Multicommodity Capacitated Location Problem with Balancing Requirements. *Parallel Computing*, 29:591–606.

Gendron, B., Potvin, J.-Y., and Soriano, P. (2003b). A Tabu Search with Slope Scaling for the Multicommodity Capacitated Location Problem with Balancing Requirements. *Annals of Operations Research*, 122:193–217.

Golden, B.L. and Assad, A.A., editors (1988). *Vehicle Routing: Methods and Studies*. North-Holland, Amsterdam.

Griffiths, J.D. (1976). Optimal Handling Capacity at a Berth. *Maritime Studies and Management*, 3:163–167.

Grünert, T. and Sebastian, H.-J. (2000). Planning Models for Long-haul Operations of Postal and Express Shipment Companies. *European Journal of Operational Research*, 122(2):289–309.

Grünert, T., Sebastian, H.-J., and Thäerigen, M. (1999). The Design of a Letter-Mail Transportation Network by Intelligent Techniques. In Sprague, R., editor, *Proceedings Hawaii International Conference on System Sciences 32.*

Grunow, M., Günther, H.-O., and Lehmann, M. (2004). Dispatching Multi-load AGVs in Highly Automated Seaport Container Terminals. *OR Spectrum*, 26(2):211–235.

Guan, Y. and Cheung, R.K. (2004). The Berth Allocation Problem: Models and Solution Methods. *OR Spectrum*, 26(1):75–92.

Guan, Y., Xiao, W.Q., Cheung, R.K., and Li, C.L. (2002). A Multiprocessor Task Scheduling Model for Berth Allocation: Heuristic and Worst-Case Analysis. *Operations Research Letters*, 30(5):343–350.

Guélat, J., Florian, M., and Crainic, T.G. (1990). A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows. *Transportation Science*, 24(1):25–39.

Günther, H.-O. and Kim, K.H., editors (2005). *Container Terminals and Automated Transport Systems*. Springer-Verlag, Berlin.

Harker, P.T. (1987). *Predicting Intercity Freight Flows*. VNU Science Press, Utrech, The Netherlands.

Harker, P.T. (1988). Issues and Models for Planning and Regulating Freight Transportation Systems. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics*, pages 374–408. Springer-Verlag, Berlin.

Harker, P.T. and Friesz, T.L. (1986a). Prediction of Intercity Freight Flows I: Theory. *Transportation Research Part B: Methodological*, 20B(2):139–153.

Harker, P.T. and Friesz, T.L. (1986b). Prediction of Intercity Freight Flows II: Mathematical Formulations. *Transportation Research Part B: Methodological*, 20B(2):155–174.

Holguin-Veras, J. and Jara-Diaz, S. (1999). Optimal Pricing for Priority Service and Space Allocation in Container Ports. *Transportation Research Part B: Methodology*, 33B:81–106.

Hurley, W.J. and Petersen, E.R. (1994). Nonlinear Tariffs and Freight Network Equilibrium. *Transportation Science*, 28(3):236–245.

Imai, A., Nagaiwa, K., and Tat, C.W. (1997). Efficient Planning of Berth Allocation for Container Terminals in Asia. *Journal of Advanced Transportation*, 31(1):75–94.

Imai, A., Nishimura, E., and Papadimitriou, S. (2001). The Dynamic Berth Allocation Problem for a Container Port. *Transportation Research Part B: Methodology*, 35B(4):401–417.

Imai, A., Nishimura, E., and Papadimitriou, S. (2003). Berth Allocation with Service Priority. *Transportation Research Part B: Methodology*, 37B(5):437–457.

Isard, W. (1951). Interregional and Regional Input-Output Analysis: A Model of a Space-Economy. *The Review of Economics and Statistics*, 33:318–328.

ISL (2005). ISL Shipping Statistics and Market review (SSMR) - Short Comment (10) 2005. http://www.isl.org.

Jaillet, P., Song, G., and Yu, G. (1996). Airline Network Design and Hub Location Problems. *Location Science*, 4(3):195–212.

Jones, P.S. and Sharp, G.P. (1977). Multi-Mode Intercity Freight Transportation Planning for Underdeveloped Regions. In *Proceedings of the 18th Meeting of the Transportation Research Forum*, pages 523–531.

Jordan, W.C. and Turnquist, M.A. (1983). A Stochastic Dynamic Network Model for Railroad Car Distribution. *Transportation Science*, 17:123–145.

Jourquin, B. and Beuthe, M. (1996). Transportation Policy Analysis with a Geographic Information System: The Virtual Network of Freight Transportation in Europe. *Transportation Research Part C: Emerging Technologies*, 4(6):359–371.

Kim, D., Barnhart, C., Ware, K., and Reinhardt, G. (1999). Multimodal Express Package Delivery: A Service Network Design Application. *Transportation Science*, 33(4):391–407.

Kim, K.H. (1997). Evaluation of the Number of Rehandles in Container Yards. *Computers & Industrial Engineering*, 32(4):701–711.

Kim, K.H. and Bae, J.W. (1998). Remarshaling Export Containers in Port Container Terminals. *Computers & Industrial Engineering*, 35(3-4):655–658.

Kim, K.H. and Bae, J.W. (1999). A Dispatching Method for Automated Guided Vehicles to Minimize Delay of Containership Operations. *International Journal of Management Science*, 5(1):1–25.

Kim, K.H. and Bae, J.W. (2004). A Look-Ahead Dispatching Method for Automated Guided Vehicles in Automated Port Container Terminals. *Transportation Science*, 38(2):224–234.

Kim, K.H., Kang, J.S., and Ryu, K.R. (2004). A Beam Search Algorithm for the Load Sequencing of Outbound Containers in Port Container Terminals. *OR Spectrum*, 26(1):93–116.

Kim, K.H. and Kim, H.B. (1999). Segregating Space Allocation Models for Container Inventories in Port Container Terminals. *International Journal of Production Economics*, 59:415–423.

Kim, K.H. and Kim, H.B. (2002). The Optimal Sizing of the Storage Space and Handling Facilities for Import Containers. *Transportation Research Part B: Methodology*, 36B:821–835.

Kim, K.H. and Kim, K.Y. (1999b). An Optimal Routing Algorithm for a Transfer Crane in Port Container Terminals. *Transportation Science*, 33(1):17–33.

Kim, K.H. and Kim, K.Y. (1999c). Routing Straddle Carriers for the Loading Operation of Containers Using a Beam Search Algorithm. *Computers & Industrial Engineering*, 36(1):109–136.

Kim, K.H., Lee, K.M., and Hwang, H. (2003). Sequencing Delivery and Receiving Operations for Yard Cranes in Port Container Terminals. *International Journal of Production Economics*, 84(3):283–292.

Kim, K.H. and Moon, K.C. (2003). Berth Scheduling by Simulated Annealing. *Transportation Research Part B: Methodology*, 37B(6):541–560.

Kim, K.H. and Park, K.T. (2003a). A Note on a Dynamic Space-Allocation Method for Outbound Containers. *European Journal of Operational Research*, 148(1):92–101.

Kim, K.H. and Park, K.T. (2003b). Dynamic Space Allocation for Temporary Storage. *International Journal of Systems Science*, 34(1):11–20.

Kim, K.H. and Park, Y.M. (2004). A Crane Scheduling Method for Port Container Terminals. *European Journal of Operational Research*, 156(3):752–768.

Kim, K.H., Park, Y.M., and Ryu, K.R. (2000). Deriving Decision Rules to Locate Export Containers in Container Yard. *European Journal of Operational Research*, 124(1):89–101.

Kim, K.Y. and Kim, K.H. (1999a). A Routing Algorithm for a Single Straddle Carrier to Load Export Containers onto a Containership. *International Journal of Production Economics*, 59:425–433.

Kim, K.Y. and Kim, K.H. (2003). Heuristic Algorithms for Routing Yard-Side Equipment for Minimizing Loading Times in Container Terminals. *Naval Research Logistics*, 50(5):498–514.

Klincewicz, J.G. (1990). Solving a Freight Transport Problem Using Facility Location Techniques. *Operations Research*, 38(1):99–109.

Klincewicz, J.G. (1991). Heuristics for the p-Hub Location Problem. *European Journal of Operational Research*, 53(1):25–37.

Klincewicz, J.G. (1992). Avoiding Local Optima in the p-Hub Location Problem Using Tabu Search and GRASP. *Annals of Operations Research*, 40:283–302.

Klincewicz, J.G. (1996). Dual Algorithm for the Uncapacitated Hub Location Problem. *Location Science*, 4(3):173–184.

Koh, Y.K. and Kim, S.C. (2001). A Study on the Impact of Extra Large-Sized Containership on the Shipping Company & Port. *Korean International Commerce*, 16(2):165–187.

Kozan, E. (1997). Comparison of Analytical and Simulation Planning Models of Sea Port Container Terminals. *Transportation Planning and Technology*, 20:235–248.

Kozan, E. (2000). Optimizing Container Transfers at Multimodal Terminals. *Mathematical and Computer Modeling*, 31:235–243.

Kozan, E. and Preston, P. (1999). Genetic Algorithms to Schedule Container Transfers at Multimodal Terminals. *International Transactions in Operational Research*, 6:311–329.

Kuby, M.J. and Gray, R.G. (1993). The Hub Network Design Problem with Stopovers and Feeders: The Case of Federal Express. *Transportation Research Part A: Policy and Practice*, 27A(1):1–12.

Labbé, M. and Louveaux, F.V. (1997). Location Problems. In Dell'Amico, M., Maffioli, F., and Martello, S., editors, *Annotated Bibliographies in Combinatorial Optimization*, pages 261–281. John Wiley & Sons, New York, NY.

Labbé, M., Peeters, D., and Thisse, J.-F. (1995). Location on Networks. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, pages 551–624. North-Holland, Amsterdam.

Lai, K.K. and Lam, K. (1994). A Study of Container Yard Equipment Allocation Strategy in Hong Kong. *International Journal of Modeling & Simulation*, 14(3):134–138.

Lai, K.K. and Leung, J. (2000). Analysis of Gate House Operations in a Container Terminal. *International Journal of Modelling and Simulation*, 20(1):89–94.

Lai, K.K. and Leung, J.W. (1996). Analysis of Yard Crane Deployment Strategies in a Container Terminal. In *Proceedings of ICC & IE Conference*, pages 1187–1190.

Lai, K.K. and Shih, K. (1992). A Study of Container Berth Allocation. *Journal of Advanced Transportation*, 26(1):45–60.

Larin, D., Crainic, T.G., Simonka, G., James-Lefebvre, L., Dufour, G., and Florian, M. (2000). *STAN User's Manual, Release 6.* INRO Consultants, Inc., Montréal, QC, Canada.

Li, C.-L., Cai, X., and Lee, C.-Y. (1998). Scheduling with Multiple-Job-on-One-Processor Pattern. *IEE Transactions on Scheduling and Logistics*, 30(5):433–445.

Lim, A. (1998). The Berth Planning Problem. *Operations Research Letters*, 22(2-3):105–110.

Macharis, C. and Bontekoning, Y.M. (2004). Opportunities for OR in Intermodal Freight Transport Research: A Review. *European Journal of Operational Research*, 153(2):400–416.

Magnanti, T.L. and Wong, R.T. (1984). Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, 18(1):1–55.

Mattfeld, D.C. and Kopfer, H. (2004). Terminal Operations Management in Vehicle Transshipment. *Transportation Research Part A: Policy and Practice*, 37A(5):435–452.

Miller, A.J. (1971). Queuing at Single-Berth Shipping Terminal. *Journal of Waterways, Harbors and Coastal Engineering Division*, 97:43–56.

Minoux, M. (1989). Network Synthesis and Optimum Network Design Problems: Models, Solution Methods and Applications. *Networks*, 19:313–360.

Mirchandani, P.S. and Francis, R.L., editors (1990). *Discrete Location Theory*. John Wiley & Sons, New York, NY.

Nagurney, A. (1993). *Network Economics – A Variational Inequality Approach*. Kluwer Academic Publishers, Norwell, MA.

Nam, K.-C., Kwak, K.-S., and Yu, M.-S. (2002). Simulation Study of Container Terminal Performance. *Journal of Waterway, Port, Coastal and Ocean Engineering*, (May/June):126–132.

Narasimhan, A. and Palekar, U.S. (2002). Analysis and Algorithms for the Transtainer Routing Problem in Container Port Operations. *Transportation Science*, 36(1):63–78.

Nemhauser, G.L. and Wolsey, L.A. (1988). *Integer and Combinatorial Optimization*. Wiley, New York, NY.

Noritake, M. and Kimura, S. (1990). Optimum Allocation and Size of Seaports. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 116(2):287–299.

O'Kelly, M.E. (1987). A Quadratic Integer Program for the Location of Interacting Hub Facilities. *European Journal of Operational Research*, 32(3):393–404.

O'Kelly, M.E. (1992a). A Clustering Approach to the Planar Hub Location Problem. *Annals of Operations Research*, 40:339–353.

O'Kelly, M.E. (1992b). Hub Facilities Location with Fixed Costs. *Papers in Regional Science*, 71:292–306.

O'Kelly, M.E., Bryan, D., Skorin-Kapov, D., and Skorin-Kapov, J. (1996). Hub Network Design with Single and Multiple Allocation: A Computational Study. *Location Science*, 4(3):125–138.

O'Kelly, M.E., Skorin-Kapov, D., and Skorin-Kapov, J. (1995). Lower Bounds for the Hub Location Problem. *Management Science*, 41(4):713–721.

Park, K.T. and Kim, K.H. (2002). Berth Scheduling for Container Terminals by Using a Subgradient Optimization Technique. *Journal of the Operational Research Society*, 53(9):1049–1054.

Park, Y.M. and Kim, K.H. (2003). A Scheduling Method for Berth and Quay Cranes. *OR Spectrum*, 25(1):1–23.

Pedersen, M.B., Crainic, T.G., and Madsen, O.B.G. (2006). A Design-Balanced Network Design Model for Transportation Services. Publication, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Peterkofsky, R.I. and Daganzo, C.F. (1990). A Branch and Bound Solution Method for the Crane Scheduling Problem. *Transportation Research Part B: Methodology*, 24B(3):159–172.

Pirkul, H. and Jayaraman, V. (1996). Production, Transportation, and Distribution Planning in a Multi-Commodity Tri-Echelon System. *Transportation Science*, 30(4):291–302.

Pirkul, H. and Jayaraman, V. (1998). A Multi-Commodity, Multi-Plant, Capacitated Facility Location Problem: Formulation and Efficient Heuristic Solution. *Computers & Operations Research*, 25(10):869–878.

Powell, W.B. (1986). A Local Improvement Heuristic for the Design of Less-than-Truckload Motor Carrier Networks. *Transportation Science*, 20(4):246–357.

Powell, W.B. (1988). A Comparative Review of Alternative Algorithms for the Dynamic Vehicle Allocation Problem. In B.L. Golden and A.A. Assad, editor, *Vehicle Routing: Methods and Studies*, pages 249–292. North-Holland, Amsterdam.

Powell, W.B. (2003). Dynamic Models of Transportation Operations. In Graves, S. and Tok, T.A.G., editors, *Supply Chain Management*, volume 11 of *Handbooks in Operations Research and Management Science*, pages 677–756. North-Holland, Amsterdam.

Powell, W.B., Bouzaïene-Ayari, B., and Simaõ, H.P. (2006). Dynamic Models for Freight Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, Handbooks in Operations Research and Management Science. North-Holland, Amsterdam. to appear.

Powell, W.B. and Carvalho, T.A. (1998). Real-Time Optimization of Containers and Flatcars for Intermodal Operations. *Transportation Science*, 32(2):110–126.

Powell, W.B., Jaillet, P., and Odoni, A. (1995). Stochastic and Dynamic Networks and Routing. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, pages 141–295. North-Holland, Amsterdam.

Powell, W.B. and Sheffi, Y. (1983). The Load-Planning Problem of Motor Carriers: Problem Description and a Proposed Solution Approach. *Transportation Research Part A: Policy and Practice*, 17(6):471–480.

Powell, W.B. and Sheffi, Y. (1986). Interactive Optimization for Motor Carrier Load Planning. *Journal of Business Logistics*, 7(2):64–90.

Powell, W.B. and Sheffi, Y. (1989). Design and Implementation of an Interactive Optimization System for the Network Design in the Motor Carrier Industry. *Operations Research*, 37(1):12–29.

Powell, W.B. and Topaloglu, H. (2003). Stochastic Programming in Transportation and Logistics. In Ruszczynski, A. and Shapiro, A., editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 555–635. North-Holland, Amsterdam.

Powell, W.B. and Topaloglu, H. (2005). Fleet Management. In Wallace, S. and Ziemba, W., editors, *Applications of Stochastic Programming*, Math Programming Society - SIAM Series on Optimization, pages 185–216. SIAM, Philadelphia, PA.

Preston, P. and Kozan, E. (2001). An Approach to Determine Storage Locations of Containers at Seaport Terminals. *Computers & Industrial Engineering*, 28:983–995.

Ramani, K.V. (1996). An Interactive Simulation Model for the Logistics Planning of Container Operations in Seaports. *Simulation*, 66(5):291–300.

Roy, J. and Crainic, T.G. (1992). Improving Intercity Freight Routing with a Tactical Planning Model. *Interfaces*, 22(3):31–44.

Roy, J. and Delorme, L. (1989). NETPLAN: A Network Optimization Model for Tactical Planning in the Less-than-Truckload Motor-Carrier Industry. *INFOR*, 27(1):22–35.

Ryu, K.R., Kim, K.H., Lee, Y.H., and Park, Y.M. (2001). Load Sequencing Algorithms for Container Ships by Using Metaheuristics. In *Proceedings of 16th International Conference on Production Research*. Prague, Czech Republic, 29 July-4 August, CD-ROM.

Salkin, H.M. and Mathur, K. (1989). *Foundations of Integer Programming*. North-Holland, Amsterdam.

Schonfeld, P. and Sharafeldien, O. (1985). Optimal Berth and Crane Combinations. *Journal of Waterway, Port, Coastal and Ocean Engineering*, 111(6):1060–1072.

Sharp, G.P. (1979). A Multi-Commodity, Intermodal Transportation Model. In *Proceedings of the 20th Meeting of the Transportation Research Forum*, pages 399–407.

Skorin-Kapov, D. and Skorin-Kapov, J. (1994). On Tabu Search for the Location of Interacting Hub Facilities. *European Journal of Operational Research*, 73(3):502–509.

Skorin-Kapov, D., Skorin-Kapov, J., and O'Kelly, M.E. (1996). Tight Linear Programming relaxation of Uncapacitated p-Hub Median Problems. *European Journal of Operational Research*, 94(3):582–593.

Smith, K., Krishnamoorthy, M., and Palaniswami, M. (1996). Neural Versus Traditional Approaches to the Location of Interacting Hub Facilities. *Location Science*, 4(3):155–171.

Steenken, D., Voß, S., and Stahlbock, R. (2004). Container Terminal Operation and Operations Research - A Classification and Literature Review. *OR Spectrum*, 26(1):3–49.

Taleb-Ibrahimi, M., Castilho, B., and Daganzo, C.F. (1993). Storage Space vs Handling Work in Container Terminals. *Transportation Research Part B: Methodology*, 27B(1):13–32.

Toth, P. and Vigo, D., editors (2002). *The Vehicle Routing Problem*. SIAM Monographs on Discrete Mathematics and Applications. SIAM, Philadelphia, PA.

van der Heyden, W.P.A. and Ottjes, J.A. (1985). A Decision Support System for the Planning of the Workload of a Grain Terminal. *Decision Support Systems*, 1:293–297.

van Hee, K.M. and Wijbrands, R.J. (1988). Decision Support System for Container Terminal Planning. *European Journal of Operational Research*, 34(3):262–272.

Vis, I.F.A., de Koster, R., and Roodbergen, K.J. (2001). Determination of the Number of Automated Guided Vehicles Required at a Semi-automated Container Terminal. *Journal of the Operational Research Society*, 52(4):409–417.

Wanhill, S.R.C. (1974). Further Analysis of Optimum Size Seaport. *Journal of the Waterways Harbors and Coastal Engineering Division*, 100:377–383.

White, W. (1972). Dynamic Transshipment Networks: An Algorithm and its Application to the Distribution of Empty Containers. *Networks*, 2(3):211–236.

Winston, C. (1983). The Demand for Freight Transportation: Models and Applications. *Transportation Research Part A: Policy and Practice*, 17:419–427.

Zhang, C., Liu, J., Wan, Y.-W., Murty, K.G., and Linn, R. (2003). Storage Space Allocation in Container Terminals. *Transportation Research Part B: Methodology*, 37B(10):883–903.

Zhang, C., Wan, Y.W., Liu, J., and Linn, R.J. (2002). Dynamic Crane Deployment in Container Storage Yards. *Transportation Research Part B: Methodology*, 36B(6):537–555.